

The background of the page features a large, light gray watermark of the Open University logo. The logo consists of a stylized '@' symbol inside a circle, with the word 'Open' to its left and 'University' below it. The text 'ignou' is also visible in a large, light gray font across the middle of the page.

BLOCK 4

INTERNET RESOURCES AND SERVICES

UNIT 12 BASICS OF INTERNET

Structure

- 12.0 Objectives
- 12.1 Introduction
- 12.2 History of Internet
- 12.3 Internet Management
- 12.4 Definitions of Internet
- 12.5 Growth of Internet
 - 12.5.1 Internet 2
 - 12.5.2 Internet 3
- 12.6 Internet Architecture
 - 12.6.1 Peer-to-Peer Communication
 - 12.6.2 Client Server Architecture
- 12.7 Accessing the Internet
 - 12.7.1 Analog Dial-up Connection
 - 12.7.2 Digital Subscribers Line (DSL)
 - 12.7.3 Dedicated Leased Line
 - 12.7.4 Integrated Services Digital Network (ISDN)
 - 12.7.5 Cable Connection to the Internet
 - 12.7.6 Internet through Mobile Telephone
 - 12.7.7 Shell v/s TCP/IP Accounts
- 12.8 Internet Service Providers (ISPs)
 - 12.8.1 Choosing ISP
 - 12.8.2 Internet Service Providers in India
- 12.9 Hardware and Software Required for the Internet
 - 12.9.1 Server-side Hardware Components
 - 12.9.2 Server-side Software Components
 - 12.9.3 Client-side Hardware and Software Components
- 12.10 Internet Standards and Protocols
 - 12.10.1 Internet Protocols
 - 12.10.2 Transmission Control Protocol / Internet Protocol (TCP / IP)
 - 12.10.3 Hypertext Transfer Protocol (HTTP)
 - 12.10.4 File Transfer Protocol (FTP)
 - 12.10.5 Serial Line Internet Protocol (SLIP) and Point-to-Point Protocol (PPP)
 - 12.10.6 Z39.50
 - 12.10.7 Dublin Core or Z39.85

- 12.11 Organisation of Internet
 - 12.11.1 Internet Addressing
 - 12.11.2 IP Addresses
 - 12.11.3 Domain Name
 - 12.11.4 Uniform Resource Locator (URL)
- 12.12 Internet Security
 - 12.12.1 User Authentication
 - 12.12.2 Fire Walls
 - 12.12.3 Proxy Servers
 - 12.12.4 Data Encryption
 - 12.12.5 Digital Certification
 - 12.12.6 Intrusion Detection System (IDS)
 - 12.12.7 Computer Viruses
 - 12.12.8 Virtual Private Network (VPN)
 - 12.12.9 Extranet
- 12.13 Summary
- 12.14 Answers to Self Check Exercises
- 12.15 Keywords
- 12.16 References and Further Reading

12.0 OBJECTIVES

After reading this Unit, you will be able to acquire knowledge on the following components of Internet:

- Internet history, growth and development and its management;
- Internet architecture, methods of accessing Internet and software and hardware requirement for Internet for accessing Internet;
- Internet standards and protocols and organisation of Internet; and
- Internet security.

12.1 INTRODUCTION

The Internet has emerged as one of the most powerful media of communication. The Internet and associated technologies have created a global environment that has transformed the world into a global village defying the limitations of geographical boundaries. About two decades ago, most of the world knew very little or nothing about the Internet. Originating from a network owned by the US Defence project, the Internet, till 1970s connected a limited number of computers accessible only to computer scientists and researchers in USA and allied countries for the purpose of defence research. Initially, it became accessible to universities and research institutions to facilitate easier and faster communication between scientists and researchers. With winding of operations of ARPANET in 1990, the combined infrastructure of ARPANET and NSFNET became available commercially. The Internet, since then, is growing at leaps and bounds around the world and broadband is rapidly becoming a large part of the growth. The number of hosts is considered an accurate measure of the size of the Internet. The Domain Survey,

sponsored by the Internet Systems Consortium that discovers every server on the Internet reports that Internet host count has reached over 233 million and has grown by 35 percent in the past year. The InternetWorldStats.com reports that there are more than 745 million Internet users worldwide, with an average of 3.2 users per host. With the high quality of service in the United States there are approximately 2.4 Internet users per host, whereas in some developing countries such as China and India, there are more than 100 Internet users per host. The Internet statistics can be used as a barometer of the global economy. For example, in case of India, the total number of users increased from 5 million in the year 2000 to 16.58 million in the beginning of 2004. This 231.6% increase in the growth of number of users in four years is an indication of Internet-based economic activity in India as well as overall economic changes in the country.

The Internet has completely revolutionised the modes and methods of the computer and communications world like never before. The Internet is an integration of several functions and services rolled into one. It is a world-wide broadcasting facility, a mechanism for information storage and dissemination, a media for electronic publishing of scholarly literature and a medium for collaboration and interaction between individuals and their computers without barriers of geographic location. The Internet represents one of the most successful examples of the benefits of sustained investment and commitment to research and development in information infrastructure. Beginning with the early research in packet switching, the government, industry and academia have been partners in evolving and deploying this exciting new technology.

12.2 HISTORY OF INTERNET

The history of Internet can be traced back to 1957 when erstwhile Soviet Union launched its first satellite, Sputnik I, prompting US President Dwight Eisenhower to launch Defence Advanced Research Projects Agency (DARPA) to regain the lead in the technological race. DARPA's mission was to advance science and technology for military applications. The DARPA developed its first successful satellite in 18 months. By the end of 1960, it began to focus on computer networking and communication technology essentially to establish communication links between research centres and universities established across the country as part of its overall mission. ARPANET was commissioned in 1969 and by 1971 it had 15 nodes and 23 hosts. The e-mail was invented in 1972 by Ray Tomlinson to send messages across a distributed network. In 1973, the first international connection to the evolving Internet was established at the University College of London and the Royal Radar Establishment (Norway). In the same year, DARPA initiated a research program to investigate techniques and technologies for interlinking packet networks of various kinds. The objective was to develop communication protocols based on 'packet-switching' that would allow networked computers to communicate seamlessly across multiple, geographically dispersed locations. The 'packet-switching' would split the data to be transmitted into tiny packets that can take different routes to their destination. Such a network would withstand partial nuclear attack feared during cold war. Moreover, interception of data travelling on the net in packets would be difficult. This was called the Internetting project and the system of networks that emerged from the research was known as the 'Internet'. The system of protocols which was developed over the course of this research effort became known as the TCP/IP Protocol Suite, after the two initial protocols developed: Transmission Control Protocol (TCP) and Internet Protocol (IP).

The operation management of the emerging Internet was handed over to Defence Communication Agency (DCA) in 1975. The Unix to Unix Copy Program (UUCP) was developed at the Bell Labs (AT & T) in the year 1976. The year 1977 witnessed the development of mail specifications (RFC 733). Usenet was established in the same year using UUCP (Unix to Unix Copy Program) between Duke and the University of North Carolina (UNC). DARPA also established the Internet Configuration Control Board (ICCB) in the year 1977.

In 1981, CSNET (Computer Science Network) was built with collaboration of a number of universities and industries in USA. The National Science Foundation gave financial support to the CSNET to provide networking services. CSNET used the Phonenet MMDF protocol for telephone-based electronic mail relaying and, in addition, pioneered the first use of TCP/IP over X.25 using commercial public data networks. The CSNET server provided an early example of white pages directory service and this software is still in use at numerous sites. At its peak, CSNET had approximately 200 participating sites and international connections to approximately fifteen countries. Another important development in the same year was the creation of BITNET (Because it's time network). The BITNET was started as a cooperative network at the City University of New York with first connection to University of Yale. BITNET adopted the IBM RSCS protocol suite that connected participating sites through leased lines. Most of the original BITNET connections linked IBM mainframes in university data centres. From the beginning, BITNET has been multi-disciplinary in nature with users in all academic areas. It has also provided a number of unique services to its users (e.g., LISTSERV). Today, BITNET and its parallel networks in other parts of the world (e.g., EARN in Europe) have several thousand participating sites. In recent years, BITNET has established a backbone, which uses the TCP/IP protocols with RSCS-based applications running above TCP.

The year 1982 was a year of great significance in the growth and development of Internet. Defence Communication Agency (DCA) and DARPA adopted Transmission Control Protocol (TCP) and Internet Protocol (IP) suite (commonly known as TCP/IP) as the official protocol suite for ARPANET. This led to one of the first definition of Internet as connected set of networks using TCP/IP. In the same year, the Eunet (European UNIX Network) was created to provide e-mail and Usenet services in Europe. The External Gateway Protocol (EGP) was also developed in the same year, which defines protocols for connecting networks that were not based on TCP/IP with the Internet. The University of Wisconsin developed 'Name Server' in 1982 that facilitated translation of names into strings of numbers. This development led to the practice of assigning domain names for the sites that is being practiced even now. Other significant development that took place in 1982 included splitting of ARPANET into ARPANET and MILNET. The MILNET was later integrated with the Defence Data Network created in 1981.

Launch of desktop computers in 1982 led to major shift from having a single, large main frame computer connected to the Internet on each site to the entire local areas network connected to the Internet. In the same year, the Internet Activities Board (IAB) replaced ICCB with a primary mission to guide evolution of the TCP / IP protocol suite and to provide research advice to the Internet community.

Domain Name Servers as distributed databases were introduced in 1984 to facilitate translation from domain names to IP addresses. Transition to naming standards from numeric addresses proved to be very helpful in popularisation of the Internet. For example, it is much easy to remember www.yahoo.com than its numerical equivalent.

In 1986, the U.S. National Science Foundation (NSF) initiated the development of the NSFNET, which today provides a major backbone communication service for the Internet. The National Aeronautics and Space Administration (NASA) and the U.S. Department of Energy contributed additional backbone facilities in the form of the NSFNET and ESNET respectively. The Network News Transfer Protocol (NNTP) was designed to enhance news performance over TCP/IP.

In 1987, the NSF signed a cooperative agreement to manage the NSFNet backbone with Merit Networks, Inc. Merit, IBM and MCI later founded Advanced Network and Services, Inc. (ANS). In the same year, BITNET and CSNET merged to form the Corporation for Research and Educational Networking (CREN). In the fall of 1991, CSNET service was discontinued having fulfilled its important early role in the provision of academic networking service. A key feature of CREN is that its operational costs were fully met through dues paid by its member organisations.

A computer virus for the first time affected approximately 6,000 of total 60,000 hosts on the Internet in the year 1988. The vulnerability of Internet and the need for more security was realised for the first time. DARPA formed the Computer Emergency Response Team (CERT) in response. In the same year, Department of Defence adopted Open Systems Interconnection (OSI).

The total number of hosts on the Internet rose to 100,000 in 1989. The year also witnessed first relays between a commercial electronic mail carrier and the Internet. MCI Mail connected through the Corporation for the National Research Initiative (CNRI) and CompuServe connected through Ohio State University. The Corporation for Research and Education Networking (CREN) was formed with the merger of CSNET and BITNET. The Internet Engineering Task Force (IETF) and Internet Research Task Force (IRTF) also came into existence under the IAB in the year 1989. In the same year, several other countries got connected to the NSFNet including Australia, Germany, Israel, Italy, Japan, Mexico, the Netherlands, New Zealand, Puerto Rico and the United Kingdom. In Europe, major international backbones such as NORDUNET and others provide connectivity to over one hundred thousand computers on a large number of networks. During the course of its evolution, particularly after 1989, the Internet system began to integrate support for other protocol suites into its basic networking architecture. The present emphasis in the system is on multi-protocol internetworking, and in particular, with the integration of the Open Systems Interconnection (OSI) protocols into the architecture.

During the early 1990's, OSI protocol implementations also became available and, by the end of 1991, the Internet has grown to include some 5,000 networks in over three dozen countries, serving over 700,000 host computers used by over 4,000,000 people. The ARPANET ceased to exist in 1990. Commercial network providers in the U.S. and Europe began to offer Internet backbone and access support on a competitive basis to interested parties. Access to Internet was first offered on commercial basis by 'World' (world.std.com), thus it became the first Internet Service Provider (ISP) of Internet dial-up access. Several other countries got connected to the Internet in 1990 including Argentina, Austria, Belgium, Brazil, Chile, Greece, India, Ireland, South Korea, Spain and Switzerland.

Wide Area Information Servers (WAISs) were invented in 1991 by Brewster Kahle and released by the Thinking Machines Corporation. These servers became the basis of indices to information available on the Internet. The indexing and search techniques implemented by these engines allow Internet users to find information using keywords across vast resources available on the net.

The most significant development in the history of Internet was the invention of World Wide Web (WWW) by Tim Berners-Lee at the CERN Laboratory in 1991. The first web browser called 'Mosaic' was released in 1993 that took the Internet by storm. Several other countries got connected to the Internet in the year 1993. The InterNIC was created in 1993 to provide specific Internet services including i) Directory of database services; ii) Registration services; and iii) Information services.

In 1994, the Internet (ARPANET) celebrated its 25th anniversary. Internet shopping and e-commerce commenced its operation on the net. Growth on the Internet traffic became geometric, i.e., NSFNet traffic passed 10 trillion bytes/month during 1994. WWW became the second most popular service on the net (behind FTP) leaving Telnet at third place. In March 1995, the WWW surpassed FTP as the service with greatest traffic on NSFNet based on packet count.

Several traditional dial-up systems in USA including CompuServe, America Online, Prodigy began to provide Internet access for services other than e-mail, i.e., WWW, Gopher, FTP and so on.

The technologies of the decade were WWW and search engines. New technologies emerged in late 1990s including client-based code loaded from web servers such as Java,

JavaScript and ActiveX, etc. The research and development on the Internet and related technologies continues even today.

A great deal of support for the Internet community has come from the U.S. Federal Government, since the Internet was originally part of a federally-funded research program and, subsequently, has become a major part of the U.S. research infrastructure. During the late 1980's, however, the population of Internet users and network constituents expanded internationally and began to include commercial facilities. Indeed, the bulk of the system today is made up of private networking facilities in educational and research institutions, businesses and in government organisations across the globe.

Self Check Exercise

- 1) What role did DARPA and NSF played in the growth and development of Internet?

Note : i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.3 INTERNET MANAGEMENT

It is often stated that the Internet does not have a central authority to control its activities. Although, true, there are several well-known organisations that work together in a relatively well-structured and roughly democratic environment to collectively participate in the research, development, and management of the Internet.

The Internet has functioned as a collaborative effort among cooperating parties. The key function of this collaborative effort is to develop and evolve specifications for TCP / IP protocol that was originally developed in the DARPA research program mentioned above. In the last five or six years, this work has been undertaken on a wider basis with support from Government agencies in many countries, industry and the academic community. The Internet Activities Board (IAB) was created in 1983 to guide the evolution of the TCP/IP Protocol Suite and to provide research advice to the Internet community.

During the course of its existence, the IAB has been reorganised several times. It now has two primary components: the Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF). The IETF is primarily responsible for further evolution of the TCP/IP protocol suite, its standardisation with the concurrence of the IAB, and the integration of other protocols into Internet operation (e.g. the Open Systems Interconnection Protocols). The Internet Research Task Force (IRTF) continues to organize and explore advanced concepts in networking under the guidance of the Internet Activities Board and with support from various government agencies.

The Internet Activities Board and Internet Engineering Task Force have a secretariat to manage its day-to-day functions. Two other functions that are critical to IAB operation are publication of documents describing the Internet and the assignment and recording of various identifiers needed for protocol operation. Throughout the development of the Internet, its protocols and other aspects of its operation have been documented first in a series of documents called Internet Experiment Notes and, later, in a series of documents

called Requests for Comment (RFCs). The latter were used initially to document the protocols of the first packet switching network developed by DARPA, the ARPANET, beginning in 1969, and have become the principal archive of information about the Internet. At present, the publication function is provided by an RFC editor.

The recording of identifiers is provided by the Internet Assigned Numbers Authority (IANA) who has delegated one part of this responsibility to an Internet Registry which acts as a central repository for Internet information and which provides central allocation of network and autonomous system identifiers, in some cases to subsidiary registries located in various countries. The Internet Registry (IR) also provides central maintenance of the Domain Name System (DNS) root database which points to subsidiary distributed DNS servers replicated throughout the Internet. The DNS distributed database is used, inter alia, to associate host and network names with their Internet addresses and is critical to the operation of the higher level TCP/IP protocols including electronic mail.

There are a number of Network Information Centers (NICs) located throughout the Internet to serve its users with documentation, guidance, advice and assistance. As the Internet continues to grow internationally, the need for high quality NIC functions increases. Although the initial community of users of the Internet were drawn from the ranks of computer science and engineering, its users now comprise of a wide range of disciplines in the sciences, arts, business, military and government administration.

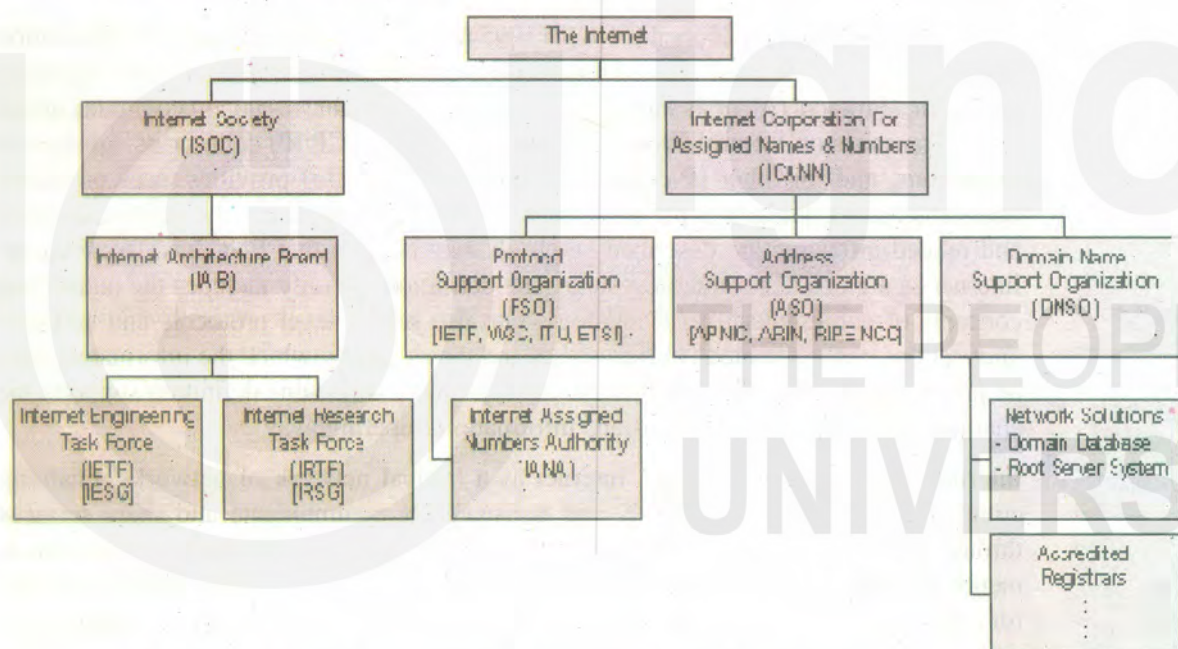


Fig. 12.1.: Internet Management (Source: www.livinginternet.com)

Self Check Exercise

2) How is Internet managed?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

.....

12.4 DEFINITIONS OF INTERNET

The term Internet has been coined from two terms, i.e., interconnection and network. A network is simply a group of computers that are connected together for sharing information and resources. Several such networks have been joined together across the world to form what is called as the Internet. The Internet is thus a network of networks. It refers to the vast collection of interconnected networks that use the TCP/IP protocols and that evolved from the ARPANET of the late 60's and early 70's (<http://1001resources.com/hosting/glossary.html>).

The Internet is the world's largest computer network that enables computers of all kinds to share services and communicate directly with each other, as if they were part of one giant seamless global computing machine. It is vast and sprawling network reaching into computer sites world-wide. The Internet comprises thousands of local area networks, groups of computers including government supercomputers, campus-wide information systems, local area networks and individual workstations. Each of these different computers, connected on Internet running on different platforms or operating systems, follows certain standards or rules of communication called protocols. The standard protocol used for Internet communication is called Transmission Control Protocol / Internet Protocol or TCP/IP. Standardised communication protocols allow similar, dissimilar, near and distant computers to communicate with one another.

The Federation National Council (FNC) in 1995 referred the Internet as Global Information System that – (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions; (ii) is able to support communications using Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions, and/ or other IP-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein. It may be seen that FNC has described the Internet as a global information system. The definition not only includes the underlying concepts of communications technology, but also higher-level protocols and end-user applications, the associated data structures and the means by which the information may be processed, manifested, or otherwise used. In many ways, this definition supports the characterisation of the Internet as an 'Information Superhighway'.

Internet Society (ISOC) defines Internet as a 'global network of networks' enabling computers of all kinds to directly and transparently communicate and share services throughout the world using a common communication protocol. It should not be seen as merely a collection of networks and computers. The Internet is an architecture that provides for both communications capabilities and information services. Because the Internet is an enormously valuable, enabling facility for so many people and organisations, it also constitutes a shared global resource of information, knowledge, and means of collaboration, and cooperation among countless diverse communities.

Self Check Exercise

- 3) Define Internet. How are different platforms or operating systems connected to the Internet?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.5 GROWTH OF INTERNET

The Internet was initially set up with 4 hosts. In early 1980s there were only 213 registered hosts on Internet. By 1986, the number had risen to 5089 hosts connected throughout the world. It was by 1989, the number of networks connected to the Internet rose to five hundred. The Network Information Centre of the Defence Data Network Information Centre found 2,218 networks connected as of January 1990. By June 1991, the National Science Foundation Network Information Centre pegged it at close to four thousand. The number of networks connected to the Internet by the end of 2003 was more than 60,000.

This high levels of connectivity has resulted in an unparalleled degree of communication, collaboration, resource sharing, and information access. The number of hosts on Internet has been doubling every year since 1981. What has driven this growth more than anything is the openness of the academic community. That openness shows up in the technology of Internet, its economics and its culture. Internet has become a forum for human communication in wide variety of disciplines ranging from computers, medicine, biosciences and social sciences etc. to art, music, sports and other recreations.

The number of hosts accessible on the Internet is considered as a fair measure of its growth. Since the early 80s, when the USA government began to share their network technology with the world, there has been growth on a scale that is hard to imagine. To put it into better perspective, the number of hosts on the Internet grew from 4 in 1969 to 213 in 1981 and 233 million in early 2004. Table 1 provides growth in number of servers from 1969 to 2004.

According to the Internet Society, a non-profit society that studies and promotes the use of the Internet, 134 countries had full Internet connection and an additional 5 countries had limited access in 1996. Surveys performed by International Data Corporation and Matrix Information and Directory Services found that as of September 1997 there were between 53 and 57 million users of the Internet worldwide. By January 1999 there were about 50 million Internet connections worldwide and have grown to 200 million users in 200 countries and territories by the year 2001.

The technologies of past two decades are WWW and search engines. While the number of web pages available on the Net reached 800 million, the number of search engines grew to a several hundreds.

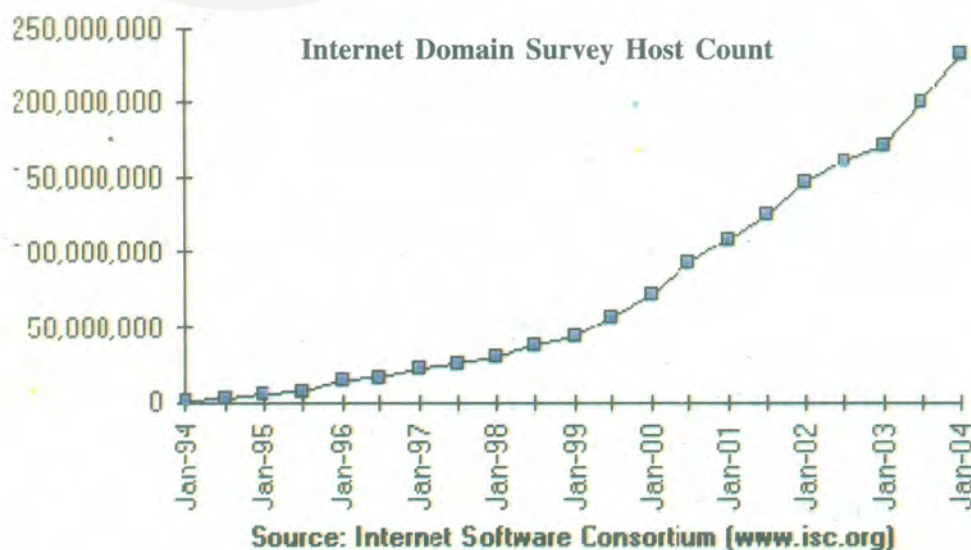


Fig. 12.2: Internet Domain Survey Host Count

Table 12.1: Total Population V/s Growth in Number of Internet Users

World Regions	Population (2004 Est.)	Internet Usage (Year 2000)	Internet Usage Latest Data	User Growth (%) (2000- 2004)	Penet- ration (%) Popu- lation)	% of Table
Africa	905,954,600	4,514,400	10,095,200	123.6	1.1	1.4
Asia	3,654,644,200	114,303,000	235,080,254	105.7	6.4	31.5
Europe	728,857,380	100,993,993	204,557,409	102.5	28.1	27.4
Middle East	259,166,000	5,272,300	14,472,500	174.5	5.6	1.9
North America	326,695,500	108,096,800	215,988,656	99.8	66.1	29.0
Latin America/ Caribbean	546,100,900	18,068,919	49,504,287	174.0	9.1	6.6
Oceania	31,892,487	7,619,500	15,654,792	105.5	49.1	2.1
India	1,088,056,200	5,000,000	16,580,000	231.6	1.5	7.2
World Total	6,453,311,067	358,871,012	745,353,098	106.3	11.5	100.0

12.5.1 Internet 2

Building on the tremendous success of Internet and associated technologies for academic needs, the university community has joined together with government and the industry to accelerate the next stage of Internet development in academia called Internet2 or I-2.

The transition from government-supported backbones to a totally privatised system in the U.S. has led to the development of a new system of backbones called Internet 2, the Next Generation Internet. Proponents of Internet2 believe that privatisation of the Internet has shifted the focus of Internet development more towards business profits and less on academics, research and teaching needs.

Internet2 is a research and development consortium led by over 206 US universities working in partnership with industry and government to develop and deploy advanced network applications and technologies, accelerating the creation of tomorrow's Internet. Internet2 is recreating the partnership among academia, industry and government that fostered today's Internet in its infancy.

In addition to university members, the Internet2 community includes over 70 companies and more than 40 affiliated organisations, including US government research laboratories. Internet2 members are working with more than 30 other similar research and education networking organisations in countries around the world. Supported by a core central staff, Internet2 activities are led by its members. Internet2 members work in concert with national, state and regional initiatives in the United States, and are coordinated with international organisations such as the Internet Engineering Task Force. Internet2 efforts are focused on the following activities:

- Advanced network applications to enable collaboration among people and provide interactive access to information and resources in ways not possible on today's commercial Internet. Interactive distance learning, remote access to unique scientific instruments, real-time access to large databases, and streaming high-definition video are all possible with high-performance networks.
- New network capabilities such as quality of service and multicasting are being aggressively tested and deployed in the networks used by Internet2 members. These capabilities support advanced network applications today, and will enable tomorrow's commercial Internet to provide the reliable performance advanced applications required.
- Middleware, the behind-the-scenes software, is providing security, directories and other services required by advanced network applications. In today's Internet, applications usually have to provide these services themselves, which leads to competing and incompatible standards. By promoting standardisation and interoperability, middleware will make advanced network applications much easier to use.

- High-performance networks are linking the campuses and laboratories of over 206 Internet member institutions. The high-performance networks participating in the Internet2 project provide the environment in which new network applications and capabilities can be deployed and tested.

The greatest benefit of Internet2 is the large bandwidth. The Internet2 is already operational. It connects 206 participating universities and institutions in US enabling transfer of large volumes of video and audio and other applications used for research purposes.

12.5.2 Internet 3

Similar to the origin of Internet, the root of the emerging Internet 3 also lies with the US government and academics. These include the USA government's Next Generation Internet (NGI) initiative (<http://www.ngi.gov/>), the National Science Foundation (NFS) and Very High Bandwidth Network Service (VBNS). As computer and communication corporate grants such as IBM, CISCO and Intel will eventually benefit with the development of Internet 3, they too are active participants in this new Internet project.

Initiated in October 1996, NGI aimed to foster partnership between academia, industry and government to develop technologies that will be essential to sustain USA's technological leadership in computing and communications and enhance the country's economic competitiveness. The NGI aimed to demonstrate the new inter-network with a capacity of 1 Terra-bps and over 10 advanced applications to leverage this bandwidth. The Internet 3 promised a large number of new applications on a very high-speed network. The Next Generation Internet (NGI) Program has been successfully completed and the Federal agencies are currently coordinating advanced networking research programs under the Large Scale Networking (LSN) Coordinating Group. The NGI Program met all of its goals except for its goal of Terabit per second networking in 2002 that is expected to meet by the current LSN research activities.

12.6 INTERNET ARCHITECTURE

The Internet uses client / server model. A server is a computer system that is accessed by other computers and / or workstations at remote locations. Usually, a server contains data, datasets, databases and programs. The server computers are also called 'host' since these computers are configured to host datasets, files and databases, receive requests for it from the client machines and serve it. The term 'host' means any computer that has full two-way access to other computers on the Internet. All computers that host web sites are host computers or servers since they 'host' information and 'serve' client machines. For example, a computer that hosts a website for 'Google' (<http://www.google.com/>) is a host or a server computer. There are millions of host computers that are linked on Internet for communicating with each other. The connectivity from one computer to another computer is being provided using some standard mode of linkages called Internet Protocols. A protocol can be defined as special set of rules governing connectivity for telecommunication connections. Protocols may exist at several levels and in order to communicate both end points must recognise and observe standard protocols. Peer to Peer and client/server are two popular systems of communication.

12.6.1 Peer to Peer Communication

Peer-to-peer is a communications model in which each party has the same capabilities and either party can initiate a communication session. In some cases, peer-to-peer communication is implemented by giving each communication node both server and client capabilities.

On the Internet, peer-to-peer (referred to as P2P) is a type of transient Internet network that allows a group of computer users with the same networking program to connect with each other and directly access files from one another's hard drives. **Napster** and **Gnutella** are examples of this kind of peer-to-peer software. Corporations are looking at the advantages of using P2P as a way for employees to share files without the expense involved in maintaining a centralized server and as a way for businesses to exchange information with each other directly. These are usually operated in small offices. IBM's

Advanced Peer-to-Peer Networking (APPN) and Gnutellanet are the examples of products that support peer-to-peer communication model.

12.6.2 Client Server Architecture

The Client-Server Architecture is based on the principle where 'client' program installed on the user's computer (called client) communicates with the 'server' program installed on the host computer to exchange information through the network. The client-server model involves two separate but related programs, i.e., client and server. The client program is loaded on the PCs of users hooked to the Internet where as the server program is loaded on to the 'host' (usually a PC with large storage capacity and RAM, a mini-computer or a main-frame computer) that may be located at a remote place. The concept of client / server computing has particular importance on the Internet because most of the programs are built using this design. A server is a program that 'serves' (or delivers) something, usually information, to a client program. A server usually runs on a computer that is connected to network. The size of that network is not important in the client / server concept – it could be a small local area network or the global Internet. The advantage of this type of design is that a server has to store the information in one format, which could be accessed by various clients working on multiple platforms, and located at different places. In the client / server model multiple client programs share the services of a common server program. Both client programs and server programs are often part of a larger program or application. For example, Internet Explorer (web browser) is a part of Windows Operating System, Internet Information Server (IIS) is a part of Windows 2000 / Windows NT and 'Apache' (web server) comes integrated with Linux Operating System. In case of World Wide Web (WWW) services, the web browser (Internet Explorer or Netscape Navigator) is a client program that resides on a PC of a user. The web browser requests services from a web server loaded on to a host machine. The 'server' program is designed to interact with 'client' programs so that a user can determine whether the information they want is available on a server, and if so, the server is programmed to serve it to the client.

Software tools in a client-server environment work in pair. For every application in client server environment, there is a client program that is responsible for facilitating users to interact with 'server' program and explore information hosted on it. The client application works as an interface between the user and the host, collecting information about the requirement of user, translating the request into agreed language of communication between the client and the server and sending it to the relevant server computer. The server program is responsible for hosting the data and accompanied programs, receiving the request from clients, finding the information requested by client and returning it to the client. A 'server' is generally programmed to organise information stored on it, create indices and search information. The responses or data sent by the server machines are received by the client machines that decode it and convert it for appropriate display on the user's machine.

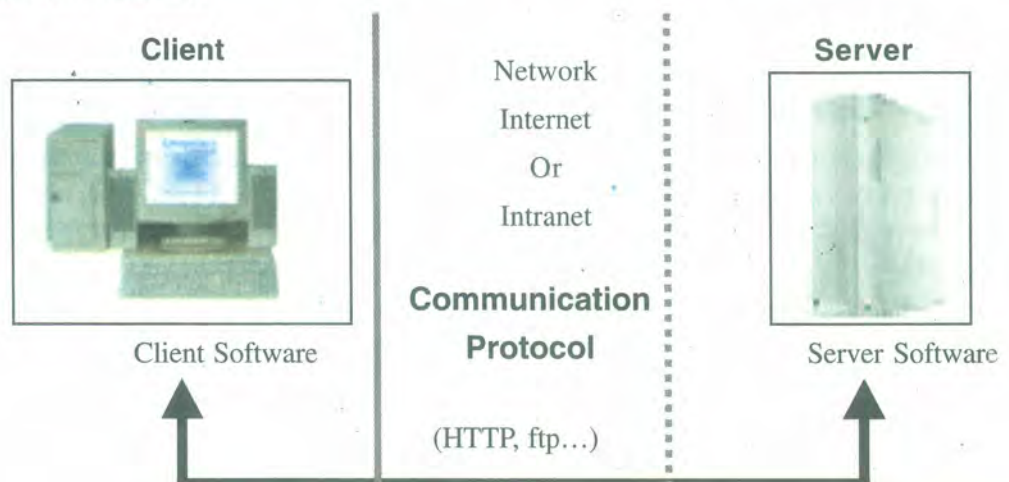


Fig. 12.3: Thematic Representation of Client / Server Model

The client-server model is most suitable for packet switching as mode of communication used by the Internet. The client-server model does not need continuous communication between client and the server. Once a request from the client machine is sent to the server that contains not only the request but also the address of client machine, the client machine can use its resources for something else while it waits for response from the server. Likewise, once a request is received from a client, the server machine performs a search for resources requested by the client and sends it to the client's address. The process does not require continuous communication between client and the server. This asynchronous form of communication is not only suitable for packet switching but it also ensures that resources of client and host machines as well as communication channels are used most optimally.

The client-server model facilitates development of different clients and servers for various applications.

The model can be extended virtually to any application. In a mailing system for example, while a mail server is configured to 'receive' and 'send' mails for mail accounts created on it for various users. The mail server program can be accessed by the users having mail accounts using mail client applications such as 'Outlook Express', 'Eudora' and 'Netscape Communicator'. The LibSys Software Package also works on client-server model. The LibSys server application resides on host computer with database of books, journals, patron, transaction records, etc. The LibSys client application resides on the machines used by the staff and the users. The client machines are configured to access OPAC for the Library users and all other modules (acquisition, technical processing, circulation, etc.) for the library staff.

Table 12.2: Functions of Server and Client Programs

Functions of a Client Program	Functions of a Server Program
- Serves as a user interface for user - Translates requests from the user into desired protocol - Sends request to the server encoded in packets - Receives response from the server - Decodes response from user and format it for presentation	- Host data and information - Organizes data and contents, develop indices to facilitates search - Receives requests from clients - Processes requests from various users - Returns results to respective clients on accompanied addresses

Table 12.3: Client-Server Programs for Various Applications

Applications	 Server Software	 Client Software
World Wide Web (WWW)	NCSA HTTPD; Apache; Jigsaw; Netra, IIS	Internet Explorer Netscape Navigator
Gopher	Gopher Server	Gopher Client
E-Mail	E-mail Server	Email Clients (Outlook Express, Eudora, Netscape Communicator)
FTP	FTP Server	FTP Client (WS_FTP, Bullet FTP)
LibSys	LibSys Server	LibSys Client

Self Check Exercise

- 4) What is client-server model? How is it different from host-terminal model?
- 5) What is packet switching? How is it different from circuit switching?

Note : i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.7 ACCESSING THE INTERNET

The most common method of accessing Internet for most users is through the telephone lines. A user dials the service provider from his / her residence or from office and the service provider puts them on to the Internet service network. This arrangement allows a user to connect to telephones almost anywhere in the world. Now that more and more service providers have started offering faster and faster options for Internet connections, users looking for faster-than-dialup connection have several options to choose from.

The connections to the Internet fall under two basic categories: dial-up access and direct or dedicated access. There are two categories of dial-up access, i.e., analogue and digital. Regular telephone lines support analogue mode of data transmission that uses a continuous wave form to transmit data. Analogue connections use modems to convert digital signals to analogue signals and then analogue signals back to the digital signals. Digital transmissions such as fibre optic devices pass data along using discrete, on / off pulses. Unlike analogue connections, digital connections do not require a modem at each end of the connection.

12.7.1 Analog Dial-up Connection

Dial-up access to the Internet is the most common way that is used by individuals and small organisations to connect to the Internet. A user connects to the Internet by dialling to the Internet service provider via a modem and a regular telephone line, i.e., Public Switched Telephone Network (PSTN). Dial-up connections to the Internet are not permanent connections.

Dial-up connections are less expensive as compared to the dedicated leased connection. It is the least expensive means of accessing the Internet. In India, dial-up connection from a typical Internet Service Provider costs anywhere between Rs. 300 to Rs. 1,000 per month for 100 to 500 hours. Dial-up connection requires very modest and low-cost hardware and software.

Although least expensive, dial-up connections to the Internet have certain disadvantages. Dial-up connection suffers from the disadvantage of low speed and less reliability. The speed of an analog dial-up connection is determined by the speed of the modem. Regular telephone lines used for accessing Internet may be slow, unreliable and busy during peak-hours. There are two types of accounts that one can establish with an analog dial-up connection: SLIP/PPP and shell accounts.

The hardware requirements common for all types of connections for user (client) and server are dealt separately. Dial-up connection to Internet specifically requires a Modem (Modulator / Demodulator).

A MODEM is a device that enables a computer to transmit data over telephone lines. Modulator converts the discrete stream of digital 'on-off' electric pulses used by the computers into the analog wave patterns used for transmission of human voice. Demodulator recovers the digital data from the transmitted analog signal. A modem can be fixed internally into PCs or it can also be bought as external device.

A reliable and high-quality Modem is a critical requirement to ensure quality and reliability of transmitted information. The Modem should incorporate error-correction protocols and should be supported by local telecommunication facilities. It should work in asynchronous mode with a speed of 33.6 to 56 Kbps.

Software Requirements

The software requirements common for all types of connections for user (client) and server are dealt separately. Dialup connection specifically requires communication software.

Communication software is a program, which establishes contact between a computer and the Internet Service Provider (ISP) using modem. Such softwares have built-in functionality of automatic dialling and automatic logon procedures including dialling to the ISP, supplying User ID and Password, etc. Communication software is built into operating systems. Windows Operating System incorporates 'Hyper Terminal' as communication software. ISPs may also provide their own communication software.

12.7.2 Digital Subscriber Line (DSL)

The Digital Subscriber Line (DSL) refers to technology that works on the existing copper telephone line, also known as a 'twisted pair'. It uses sophisticated digital coding to utilise more of the existing space on the wire, without interfering with the normal phone conversations. This technology dramatically increases the digital capacity of ordinary telephone lines. It skips the analog- digital-analog conversions and sends data directly in digital format. New equipment must be installed at the phone company's office as well as at the user's home or office. The DSL speeds are tied to the distance between the customer and the telecommunication central office. The user generally must be located within 3.5 miles of a phone company switching office. The DSL offers data download rates up to 6 Mbps, depending on the variant supplied and distance from the nearest DSL service provider's exchange. Besides, Internet, the DSL-enabled lines are capable of supporting services such as fast Internet access and video or TV on demand. For DSL subscribers through the MTNL, charges for Internet access are based on hourly usage or data transfer.

The DSL technology has several variants such as ADSL, SDSL, VDSL, HDSL, SHDSL and RADSL. A brief description of these variants are given below:

- **Asymmetric DSL (ADSL):** This technology facilitates use of the whole bandwidth of a standard telephone copper cabling. It allows a subscriber to receive data (download) at speeds of up to 1.544 Megabits per second, and to send (upload) data at speeds of 128 kilobits per second. Thus, the speeds of upload and download are 'asymmetric'. While the POTS (Plain Old Telephone Service) uses the frequencies between 300 and 3100Hz, the higher frequencies normally remain unused. ADSL uses the frequencies between 30KHz and 1.1 MHz to transport data, leaving the telephone connection, as crystal and pure as ever. ADSL can be easily and seamlessly combined with the existing ISDN lines.

- **Symmetric DSL (SDSL):** This connection, used mainly by small businesses, does not allow simultaneous use of telephone at the same time, but the speed of receiving and sending data is the same.
- **VDSL (Very high bit-rate DSL):** This is a fast connection, but works only over a short distance.
- **HDSL (High data rate DSL):** HDSL is the earliest variation of DSL. The main characteristic of HDSL is that it is symmetrical or in other words an equal amount of bandwidth is available in both directions.
- **SHDSL (Single-pair High-speed Digital Subscriber Line):** SHDSL was developed as a meltdown of several symmetric DSL technologies (HDSL, SDSL, HDSL-2) producing as a result one single internationally recognised industry standard. Unlike ADSL, which is fine-tuned to the needs of a home user and can download enormous amounts of data but has limitation when large amount of data is to be uploaded. SHDSL has no problems with 'two-way' Internet traffic. With the transfer rates of up to 2.3 Mbps, this makes it a perfect high-speed solution for medium to big enterprises, branch offices, as well as high-end residential users.
- **Rate Adaptive DSL (RADSL):** This is a variation of ADSL, but the modem can adjust the speed of the connection depending on the length and quality of the line.
- **ISDL (ISDN Digital Subscriber Line):** ISDL provides up to 144-Kbps transfer rates in each direction and can be provisioned on any ISDN capable phone line. Unlike ADSL and other DSL technologies, ISDL can be deployed regardless of the distance the user is from the central office.

Some of the specific advantages of DSL connections are:

- Internet connection and telephone work simultaneously.
- The speed is much higher than a regular modem (1.5 Mbps V/s 56 Kbps).
- DSL does not necessarily require new wiring; it can work on the existing phone lines.
- The companies that offer DSL usually provide the modem as part of the installation.

Some of the disadvantages of DSL connections are:

- A DSL connection works better when it is closer to the provider's central office.
- The connection is faster for receiving data than it is for sending data over the Internet.
- The DSL is available only in metropolitan cities.

Hardware Requirements

The hardware requirements common for all types of connections for user (client) and server are dealt separately later. A DSL connection to Internet specifically requires DSL modem or network terminators.

DSL modems or network terminators are digital devices that are used to connect a computer or network to a larger network via telephone wiring using DSL techniques. Modem is a misnomer in this case since there is no conversion from digital to analog. The DSL technologies use sophisticated modulation schemes to pack data onto copper wires. They are sometimes referred to as last-mile technologies because they are used only for connections from a telephone switching station to a home or office, not between switching stations.

Most DSL devices connect to USB-port on desktop or notebook computer. It does not require any additional network interface card. Most DSL devices support multiple operating systems.



Fig.12.4: DSL Modem or Network Terminators

Software Requirements

The software requirements common for all types of connections for user (client) and server are dealt separately later. DSL connection to Internet specifically requires DSL Installation Software. However, since both telephone and ADSL service are simultaneously available from the same copper pair, a central splitter or distributed filters is required at the exchange for decoupling ADSL and telephone signals.

Since the Internet connection in DSL is available virtually, the DSL connection does not require a communication package.

12.7.3 Dedicated Leased Line

A leased circuit is a dedicated link provided between two fixed locations for exclusive use by the customer. A leased circuit may be a speech circuit, a data circuit or a telegraph circuit. Dedicated leased line can be achieved by twisted pair of telephone lines, through VSAT or radio link. Leased line charges are generally uniform for all cases and are same as applicable for point-to-point leased circuits.

Dedicated leased lines have many advantages. The major advantage is the high speed and better reliability. With a dedicated single leased line, an organisation can have many users of a local area network connected to the Internet. Being a dedicated leased line, users need not dial to connect to the Internet. All computers on a local area network using a dedicated leased line are always connected to the Internet. This type of connection is appropriate for organisations that transfer large amounts of data and have many users and workstations that must be connected to the Internet. This option requires that dedicated lines be leased through a network provider (such as Department of Telecommunication or VSNL in India) and special network hardware be installed on site, making this a complicated operation. The only disadvantage of dedicated leased line is high cost of communication and difficulties involved in its maintenance.

Hardware and Software Requirements

The hardware requirements common for all types of connections for user (client) and server are dealt separately later in section 12.9. Besides a dedicated leased line, no other hardware or software is required except PCs and local area network along with switches or routers in the organisation for its effective usage. Dedicated connection through radio link requires radio antenna installed at the user's premises. Dedicated telephone line in India is provided by the MTNL or Department of Telecommunication. MTNL is also an Internet Service Provider.

12.7.4 Integrated Services Digital Network (ISDN)

Integrated Services Digital Network (ISDN) is a state-of-the-art Public Switched Digital Network that facilitates voice, data and image transmission over the telephone line through the telephone network. ISDN is a set of protocols defining how data is transmitted over digital networks. Unlike dial-up analog service, ISDN offers a higher bandwidth and is

capable of transmitting voice and data simultaneously on the same connection. ISDN is currently the fastest and most reliable method of connecting small computers to the Internet. Complete end-to-end solutions are now available at a very affordable price. While the data transfer rates of fastest modems that can connect to the Internet are typically 14,400 or maximum 56,000 bits per second (bps) or baud, the ISDN allows up to 128,000 bps. ISDN service can be delivered over the same two copper wires that provide telephone service. Therefore, no additional wiring is required. ISDN service is available by and large in all major cities of India. Moreover, it has overseas connectivity with most of the countries. Major benefits of ISDN are as follows:

- Single connection can support voice, data and images. An ISDN subscriber can establish two simultaneous independent calls which could be voice, data, image or combination of any two whereas only one call is possible on ordinary telephone lines.
- High quality services being digital right from premises of subscribers (end to end) are available.
- Eight terminals can be connected on a single line.
- High-speed data transfer from PC to PC is possible – 64 Kbps against existing 9.6 Kbps (6 times faster).
- The call set time is very short (1-2 seconds).
- ISDN supports a number of supplementary services.

Hardware Requirements

The hardware requirements common for all types of connections for user (client) and server are dealt separately later. Hardware required for an ISDN connection to Internet are given below:

The ISDN telephone line is terminated on a common box called the Network Termination (NT) that is installed at the subscriber's premises. The Network Termination unit along with accessories is generally provided by the Internet Service Provider or can be procured by the subscriber. The terminal equipment has to be procured by the subscriber.

ISDN supports voice, data and image transmission over the telephone line. As such, its application is not restricted to Internet access. The hardware required for an ISDN connection would therefore depend on the applications that a user wants to run on an ISDN connection. Some of the ISDN equipment may be:

- ISDN feature phone: this is a simplest type of ISDN phone which has an LCD display and some additional keys
- Terminal adapter
- PC add-on ISDN card
- Video phone
- G4 fax

12.7.5 Cable Connection to the Internet

Internet can also be accessed through the cable that brings TV channels to our homes using a cable modem. Cable modem separates digital and analog signals travelling on the cable and enables a PC to send and receive data using coaxial cable on the same frequencies used by the cable TV channels. The primary advantage of cable is the large bandwidth it offers. ISDN and DSL cannot match the bandwidth and speed offered by the cable. PSTN (dial-up) offers maximum of 56 kbps, ISDN offers a maximum of 128 Kbps whereas cable starts with 64 kbps and can go upto 38 Mbps. Internet through cable is, therefore, suitable for downloading multimedia such as movies, animation, etc.

The Internet through cable uses the existing cable that already connects TV in our home. This cable is capable of carrying both video signals and data signals at the same time. Neither television nor data signals are affected by simultaneous use. Internet through cable is right solution for bandwidth intensive applications and corporate customers. Though cable has immense potentials as an access technology, it has a long way to go in terms of infrastructure, quality of service, cost factor, security, etc. Siti Cable, Ice Network Pvt. Ltd., Innomedia Technologies Pvt. Ltd., Mantra Online, BPL, Aptech Internet Ltd. and Zeenext offer Internet through cable TV in India.

Hardware Requirements

The hardware requirements common for all types of connections for user (client) and server are dealt separately later. Cable connection to Internet specifically requires a cable modem.

A cable modem is an external device that hooks up to the PC. It interacts with a Cable Modem Termination System (CMTS) installed at a central location. Cable modems use various technologies like TDMA based DOCSIS standard or more robust and modern SCDMA based TERAYON proprietary technology.

Installing cable TV also involves modification and upgradation of existing cable TV network to handle two-way data. The process involves adding signal amplifiers and coaxial cable by the local cable provider.

12.7.6 Internet through Mobile Telephone

Internet can also be accessed through mobile telephone where the mobile functions as a high-speed modem to connect a user's mobile, PC or laptop to the Internet. Reliance India Mobile uses Code Division Multiple Access (CDMA) technology with superior voice quality and high-speed data capabilities. It offers speeds up to 144 Kbps. A data cable called 'R Connect Data Cable' is required to connect Reliance Mobile to the PC. The Data Cable is to be plugged into PC and the mobile. Software has to be installed in the PC for activating the Internet connectivity. Before start using R Connect, a user is required to subscribe to the service and activate his or her phone for R Connect.

Other mobile telephone services in India like Tata Mobile also offer such features.

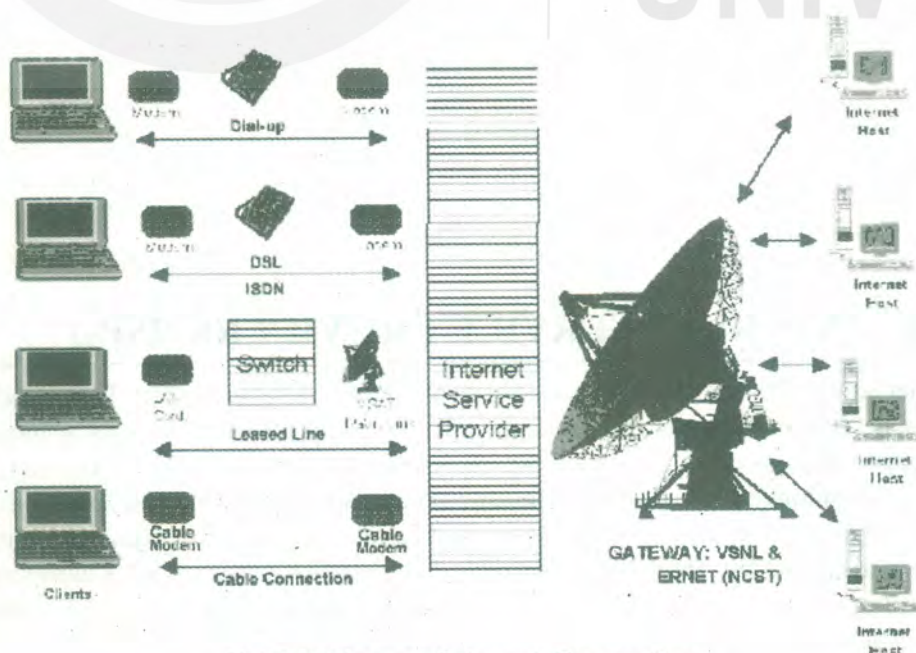


Fig. 12.5: Various Methods of Internet Connectivity

12.7.7 Shell V/s TCP/IP Accounts

A user requires an account with an Internet Service Provider (ISP) in order to access Internet services. The Internet account can either be a shell account or a TCP / IP account.

Shell Account

In case of shell account, a user logs on to an intermediary computer (host) to access the Internet. The host computer that is connected to the Internet provides connectivity to the user. There may be several shell accounts on a host computer. The primary disadvantage of a shell account is that it limits access to the Internet applications running on the service provider's computer. Moreover, a user has to learn to use the operating commands on the host computer. Moreover, shell accounts support only text-based access to the Internet.

Transferring information to or from the Internet using shell account, is a two-step process. In the first step, the file is transferred from a remote machine on the Internet to the host machine and then from host machine to the personal computer of user in the second step. Shell accounts are cheaper as compared to the TCP / IP accounts. With cheaper availability of Internet connections and popularity of graphic browsers, most users do not prefer shell accounts.

TCP / IP account

The TCP / IP account facilitates a user to configure his system as a host machine. It supports graphical interface for surfing the net. TCP / IP accounts cost more than plain shell accounts.

Self Check Exercise

- 6) Describe various methods of accessing Internet. What is the most common method of accessing Internet?
- 7) What is the limitation of Digital Subscribers Line (DSL)?

Note : i) Write your answers in the space given below.
ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.8 INTERNET SERVICE PROVIDERS (ISPs)

Internet Service Providers are companies that provide access to the Internet. The preceding section describes options for communication that a user can choose for getting an Internet connection. Various charging levels exist, but a popular method for home users is flat rate (per month unlimited time and data amount). Traditionally, the Internet was purely a text based global pool of information and access was either limited or required a certain specialised knowledge. The development of the Internet today has ensured that information now comes in other formats such as graphical, audio and animated images, and the interface for such information is now a lot more dynamic and user friendly.

To get on to the Internet, a user needs to establish a connection through an Internet Service Provider (ISP). Individuals connect their computers through the above-mentioned

communication lines through Internet service providers for accessing Internet. An ISP can be based in a small town, city or metropolitan. The small Internet Service Providers (ISPs) hook into regional ISPs, which themselves link major backbones that traverse the whole country connecting major metropolitan areas. While local ISPs may offer services only within their country, regional providers often span across state lines.

An Internet Service Provider (ISP) gives software, specialised hardware and technical help to connect to the Internet. Many ISPs provide electronic-mail account, host customers' Web pages, and offer other services as a package deal to their customers.

12.8.1 Choosing ISP

There is no reliable way fit for all to choose the best Internet Service Provider. Different people have different priorities: for some it is price, for others it is range of access numbers, for others it is speed. A user has to carefully examine which ISP addresses his or her priority in a better way.

12.8.2 Internet Service Providers (ISPs) in India

There are about 390 Internet Service Providers (ISP) in India who have been issued license by the Department of Telecommunication to set-up their ISP and tariffs. Some of them even have their own International Gateways. About 200 ISPs are already providing service to about 2.5 million users around the country. A list of selected major ISPs along with their Web site address is given below. The Videsh Sanchar Nigam Ltd (VSNL), Mahanagar Telephone Nigam Ltd. and ERNET Society are amongst the biggest Internet Service Providers in India.

Table 12.4: Selected National ISPs of India

Company	Web Page Address
Bharti Broadband Ltd.	www.mantraonline.com
BSES Telecom	www.powersufer.com
BSNL	www.bsnl.co.in , www.sancharnet.in
Caltiger	www.caltiger.com
Cyberwave	www.cyberwaveindia.com
Data Access India Ltd.	www.now-india.com
Dishnet DSL Limited	www.ddsl.net
Dishnet DSL	www.eth.net
Freedialin	www.freedialin
Glid�	www.glidemart.com
Global Online	www.gol.net.in
HCL InfiNet Ltd.	www.hclinfinet.com
Hughes Escorts	www.hughes-ecom.com
Kushagra Online	www.kushtel.net.in
Mantra Online	www.mantraonline.com
Millennium Internet	www.mol.net.in
MTNL	www.mtnl.net.in
Nanda Netcom	www.net4india.com
Nets4India	www.net4india.com
Only Smart	www.onlysmart.com
Opto Connect	www.optoconnect.net.in

Pacific Internet	www.pacific.net.in
Pioneer Online	www.pol.net.in
Pionet Onlien	www.pionetindia.com
Reliance Infocom Ltd.	www.onlysmart.com
Roltanet	www.roltanet.com
Sampark Online	www.samparkonline.com
Satyam Online	www.satyamonline.com
Sigma Online	www.sigmaonline.com
Southern Online	www.sol.net.in
Sify Ltd.	www.sifycorp.com
Spectra Net limited	www.spectranet.com
Software Tech.Parks	www.stpi.soft.net
SITI Cable Network Ltd.	www.zeenext.com
Tata Internet	www.tatanova.com
VSNL	www.vsnl.net.in
W3C	www.w3c.com
Weikfield	www.wmi.net.in
ZeeNext	www.zeenet.com

12.9 HARDWARE AND SOFTWARE REQUIRED FOR THE INTERNET

12.9.1 Server-side Hardware Components

Servers

Servers are the heart of Internet and services offered on the Internet. Server machines need to be computationally powerful, have adequate main memory (RAM) to handle requests from client machines, have large amount of secure disc storage for the database(s) and related files. It is important that the server is scalable (such as Sun Enterprise Server) so that additional storage, processing power or networking capabilities can be added wherever required.

Input Devices

Image-based web sites require input devices like scanners, digital cameras, video cameras and PhotoCD systems. A large range of choices are available for these image capturing devices. Scanners are available in all sizes and shapes. Flatbed scanners or digital cameras mounted on book cradle are more suitable for libraries.

Storage Devices

Servers hosting large web sites may require large amounts of storage; particular attention needs to be given to the storage solution. Intelligent storage networks and snapservers are now available in which the physical storage devices are intelligently controlled and made available to a number of servers. Although harddisc (fixed and removable) solutions are increasingly available at an affordable cost, optical storage devices including WORM, CD-R, CD ROM, DVD ROM or opto-magnetic devices in standalone or networked mode, are attractive alternatives for long-term storage of digital information. Optical drives record information by writing data onto the disc with a laser beam. The media offers enormous storage capabilities. A number of RAID (Redundant Array of Inexpensive

Disks) models are also available for greater security and performance. The RAID technology distributes the data across a number of disks in a way that even if one or more disks fail, the system would still function while the failed component is replaced.

12.9.2 Server-side Software Components

Setting-up a Web site may require a number of softwares which may be obtained as a single integrated package from a single vendor or it may be a system with components added onto an open architecture framework.

Image Capturing or Scanning Software

The process of converting a paper document into a computer-processible digital image is done using a software variably called document imaging system, electronic filing system or document management system, etc. A simple scanning software also comes with the scanners. Two important document imaging software from India are:

OmniDoc ver.1 (Newgen Software) <http://www.newgensoft.com/>

Data Scan (Stacks Software Pvt. Ltd.) <http://www.stex.com/>

Image Enhancement and Manipulation

The captured images may need manipulation to enhance their quality. Some of the image enhancement features include: filters, tonal reproduction, colour management, touch, crop, image sharpening, contrast, transparent background, etc. A few important image enhancement packages are listed below:

Adobe's Photoshop 9.0	http://www.adobe.com/
Jasc Inc.'s Paintshop Pro 6.02	http://www.jasc.com/
Eastman Software, Inc.	http://www.eastmansoftware.com/
Corel Corporation	http://www.corel.com/
Alchemy Mindworks	http://www.alchemy.com/

Web Servers

Setting-up a web server requires a web server program. Many server programs are available for different platforms, each with different features and cost varying from free to very expensive. Some of the important web server programs are listed below:

Servers for Unix Systems

NCSA HTTPD	http://www.ncsa.uiuc.edu/
Apache	http://www.apache.org/
Jigsaw 2.1.1	http://www.w3.org/jigsaw/
Netra (for Sun Solaris)	http://sun.com/

Servers for Windows NT

Internet Information Server (IIS)	http://www.microsoft.com/iis/
-----------------------------------	---

Database Management Software

The database management software provide structured storage and retrieval facilities to the contents of a web site. A variety of database management systems ranging from relational and extended relational database management systems to object-oriented database systems can be used for a database-driven web site. Relational DBMS are most often used for the storage of metadata and indices with attributes that contain

pointers to files in a file system. The relational DBMS software listed below can be accessed by using SQL (Structured Query Language):

Oracle	http://www.oracle.com/
Informix	http://www.informix.com/
Sybase	http://www.sybase.com/
SQL Server	http://www.microsoft.com/

12.9.3 Client-side Hardware and Software Components

A user needs a computer with necessary hardware and software devices to connect to the Internet. In order to have proper connectivity one needs a right mix of hardware and software. Depending on the need one can select hardware based on Pentium processor (Intel based) or Macintosh, preferably with latest configuration equipped with sufficient RAM, video memory, hard disc and a colour monitor, which is a basic requirement for accessing Internet.

These computers can further run on Windows, Unix, Linux or such other operating systems. All these popular operating systems now have built-in support for connecting to the Internet. In order to access the data from server computers a large number of client software are available to suite various operating systems.

The hardware devices attached to client computer also play a role in providing proper Internet connectivity. These can be either a modem or network connection. In order to connect on a Local Area Network, normally used in offices or universities/ colleges, there is a need to have Network Interface Card (NIC). These cards are designed to handle different speeds and network architecture. In order to connect from home or a small office a modem is connected to a computer. A modem can be an external device or fitted inside the computer i.e internal modem. The modems come with different speeds i.e., 14,400 bits per second (bps), 28,800 bps or 56,000 bps, etc. A modem provides connectivity to external world through various types of communication lines.

A user also requires multimedia PC (or Macintosh) equipped with an Internet Browser like Internet Explorer or Netscape Navigator to access Internet and its services. Web browsers are computer programs that facilitate a user to access the World Wide Web. They provide a graphical interface that facilitates users to click buttons, icons, and menu options to view and navigate web pages. By using web browsers, a user can locate servers of Internet, send a query, process the query results, and display them. Web browser, as a client application, is designed for a particular computing platform (for example, Windows, Macintosh, UNIX) to take advantage of the strengths of the platform. Netscape Navigator and Microsoft Internet Explorer are popular Web browsers. The client-side PCs may also require the following software packages (plug-ins) to download format-specific deliverables from the Internet:

Table 12.5: Format Specific Deliverables

Software	Used For	Web Site
Internet Explorer 5.0	Internet Browser	http://www.microsoft.com/
Netscape Navigator 4.7	Internet Browser	http://home.netscape.com/
Acrobat Reader 4.0 (Adobe)	PDF files	http://www.adobe.com/
Microsoft Office	For display and printing of MS Word, MS Access, MS Excel files and Power Point presentations	http://www.microsoft.com/
Real Player 7.0	Audio and Video	http://www.real.com/
TIFF Surfer 1.0	TIFF Images	http://www.visionshape.com/
WS_FTP Pro 6.0	File Transfer Client	http://www.ipswitch.com/

12.10 INTERNET STANDARDS AND PROTOCOLS

12.10.1 Internet Protocols

A protocol can be defined as formal description of formats and rules that two or more computers must follow to exchange data. These can be low-level details of computer-to-computer interfaces (for example the order in which bits from a byte are set across a wire) or high-level exchange between application programmes (for example the way in which two programs transfer a file across network). In simple terms, protocols are set of technical specifications that let computers exchange information.

Internet protocols are sets of rules or standard procedures that are followed to interconnect and communicate between computers in a network. The Internet Protocol allows dissimilar hosts to connect to each other through the Internet and transport information across the Internet in packets of data using the Transmission Control Protocol (TCP). This protocol also determines how to move messages and handle errors. The protocol allows creation of standards independent of hardware system. The data on Internet is transmitted from one computer to another using standard protocols.

12.10.2 Transmission Control Protocol / Internet Protocol (TCP/IP)

The two protocols pre-dominantly used by Internet are Transmission Control Protocol (TCP) and Internet Protocol (IP) and are popularly referred to as TCP/IP. These Protocols are so common for Internet that the Internet itself is defined as a "TCP/IP bound network of networks to access resources from one computer to another." These protocols were developed in 1974 by Robert Kahn of ARPANET and computer scientist Vinton G. Gerf. The greatest strength of TCP/IP is that it easily enables computers of different architectures and operating systems to communicate with each other.

TCP/IP is a two-layer program. The higher layer, Transmission Control Protocol, manages the disassembling of a message or file into smaller packets that are transmitted over the Internet and received by a TCP layer that reassembles the packets into the original message. The lower layer, Internet Protocol, handles the address part of each packet so that it gets to the right destination. Each gateway computer on the network checks this address to see where to forward the message. Even though some packets from the same message are routed differently than others, they will be reassembled at the destination.

TCP/IP uses the client/server model of communication in which a computer user (a client) requests and is provided a service (such as sending a web page) by another computer (a server) in the network. TCP/IP communication is primarily point-to-point, meaning each communication is from one point (or host computer) in the network to another point or host computer. TCP/IP and the higher-level applications that use it are collectively said to be 'stateless' because each client request is considered a new request unrelated to any previous one (unlike ordinary phone conversations that require a dedicated connection for the call duration). Being stateless free network paths, everyone can use them continuously. (Note that the TCP layer itself is not stateless as far as any one message is concerned. Its connection remains in place until all packets in a message have been received.)

The Internet Protocol (IP) is the method or protocol by which data is sent from one computer to another on the Internet. Each computer (known as a host) on the Internet has at least one IP address that uniquely identifies it from all other computers on the Internet. When a user sends or receives data (for example, an e-mail note or a web page), the message gets divided into little chunks called packets. Each of these packets contains both the sender's Internet address and the receiver's address. Packets are sent first to a gateway computer that understands a small part of the Internet. The gateway computer reads the destination address and forwards the packet to an adjacent gateway that in turn reads the destination address and so forth across the Internet until one gateway

recognizes the packet as belonging to a computer within its immediate neighbourhood or domain. That gateway then forwards the packet directly to the computer whose address is specified.

Because a message is divided into a number of packets, each packet can, if necessary, be sent by a different route across the Internet. Packets can arrive in a different order than the order they were sent in. The Internet Protocol just delivers them. It's up to another protocol, the Transmission Control Protocol (TCP) to put them back in the right order.

IP is a connectionless protocol, which means that there is no continuing connection between the two end points that are communicating. Each packet that travels through the Internet is treated as an independent unit of data without any relation to any other unit of data. (The packets do get put in the right order because of TCP, the connection-oriented protocol that keeps track of the packet sequence in a message.) In the Open Systems Interconnection (OSI) communication model, IP is in layer 3, the Networking Layer.

Another advantage of TCP/IP is that it is not bound in any way to the physical medium. Whether its wireless, token-ring, ordinary phone lines LAN or other network, one can transmit data using TCP/IP.

12.10.3 Hypertext Transfer Protocol (HTTP)

The Hypertext Transfer Protocol (HTTP) is a set of rules for exchanging files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web. As its name implies, essential concept of HTTP is the idea that files can contain links or references to other files whose selection would lead to transfer of requests from one file to another. Any web server machine contains, in addition to the HTML and other files it can serve, an HTTP daemon, a program that is designed to wait for HTTP requests and handle them when they arrive. The web browser is an HTTP client, sending requests to server machines. When a user requests for a file through browser by either 'opening' a web file (typing in a Uniform Resource Locator) or clicking on a hypertext link, the browser builds an HTTP request and sends it to the Internet Protocol address indicated by the URL. The HTTP daemon in the destination server machine receives the request and, after any necessary processing, the requested file is transmitted.

The Hypertext Transfer Protocol (HTTP) is an application-level protocol designed for a distributed, collaborative, hypermedia information system. HTTP is a client-server protocol by which two machines can communicate over a TCP/IP connection. An HTTP server is a program that waits for requests to come on a machine's port dedicated to HTTP. An HTTP client (web browser) opens a TCP/IP connection to the server via a socket, transmits a request for a document, and then waits for a reply from the server. HTTP protocol is used for document exchange in the World Wide Web. Everything that happens on the web happens over HTTP transactions.

A web page (also called a document) consists of objects. An object may be a HTML file, an image, a Java applet, a video clip, etc. that is addressable by a single URL. Most web pages consist of HTML files and several referenced objects or links.

HTTP defines how browsers request web pages from servers and how servers transfer web pages to clients i.e., in essence it defines the interaction between the web client and the web server. When a user requests a web page (for example, clicks on a hyperlink), the browser sends HTTP request messages for the objects in the page to the server. The server receives the requests and responds with HTTP response messages that contain the objects.



Fig. 12.6: HTTP Protocol Used in the Browser to Browse Web Pages

Self Check Exercise

- 8) What are protocols? Why are protocols important for functioning of Internet?
- 9) What is TCP / IP? How is it important to Internet?
- 10) Describe HTTP protocol. How does it work?

Note : i) Write your answers in the space given below.
 ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.10.4 File Transfer Protocol (FTP)

FTP is a standard Internet protocol. It is the simplest way to exchange files between computers on the Internet. Like the Hypertext Transfer Protocol (HTTP), which transfers displayable web pages and related files, and the Simple Mail Transfer Protocol (SMTP), which transfers e-mail, FTP is an application protocol that uses the Internet's TCP/IP protocols. FTP is commonly used to transfer web page files from their creator to the computer that acts as their server for everyone on the Internet. It's also commonly used to download programs and other files to client machines from servers.

As a user, one can use FTP with a simple command line interface (for example, from the Windows, MS-DOS Prompt window) or with a commercial program that offers a graphical user interface. The web browser can also make FTP requests to download programs selected from a web page. Using FTP, one can also update (delete, rename, move and copy) files at a server. Normally, one would be required to logon to an FTP server. However, publicly available files are easily accessed using anonymous FTP.

Basic FTP support is usually provided as part of a suite of programs that come with TCP/IP. However, any FTP client program with a graphical user interface usually must be downloaded from the company that makes it.

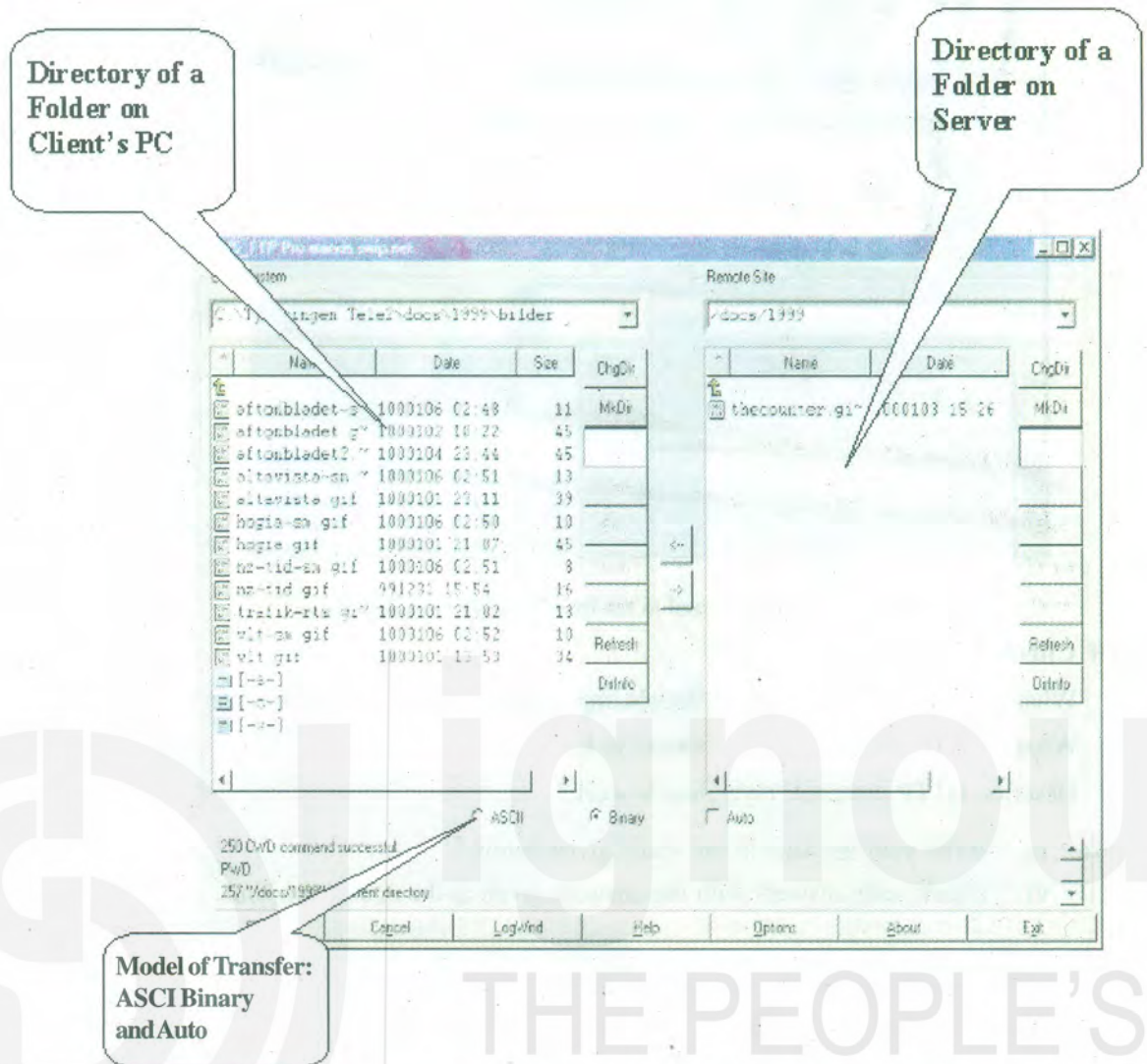


Fig.12.7: WS_FTP Client:Window Client used for Transferring Files from Client to Server and from Server to Client

12.10.5 Serial Line Internet Protocol (SLIP) and Point-to-Point Protocol (PPP)

Personal computer users usually get to the Internet through the Serial Line Internet Protocol (SLIP) or the Point-to-Point Protocol (PPP). These protocols encapsulate the IP packets so that they can be sent over a dial-up phone connection to an access provider's modem.

SLIP is a TCP/IP protocol used for communication between two machines that are previously configured for communication with each other. For example, your Internet server provider may provide you with a SLIP connection so that the provider's server can respond to your requests, pass them on to the Internet, and forward your requested Internet responses back to you. Your dial-up connection to the server is typically on a slower serial line rather than on the parallel or multiplex lines such as a line of the network you are hooking up to.

PPP (Point-to-Point Protocol) is a protocol for communication between two computers using a serial interface, typically a personal computer connected by phone line to a server. For example, your Internet server provider may provide you with a PPP connection so that the provider's server can respond to your requests, pass them on to the Internet, and forward your requested Internet responses back to you. PPP uses the Internet protocol (IP) (and is designed to handle others). It is sometimes considered a member of the TCP/IP suite of protocols. Relative to the Open Systems Interconnection (OSI) reference model,

PPP provides layer 2 (data-link layer) service. Essentially, it packages your computer's TCP/IP packets and forwards them to the server where they can actually be put on the Internet. PPP is a full-duplex protocol that can be used on various physical media, including twisted pair or fibre optic lines or satellite transmission. It uses a variation of High Speed Data Link Control (HDLC) for packet encapsulation.

PPP is usually preferred over the earlier de facto standard Serial Line Internet Protocol (SLIP) because it can handle synchronous as well as asynchronous communication. PPP can share a line with other users and it has error detection that SLIP lacks. Where a choice is possible, PPP is preferred.

12.10.6 Z39.50

Z39.50 is an American National Standard for Information Retrieval (IR). Prepared by the National Information Standards Organisation (NISO), Z39.50 defines how one system can cooperate with other systems for the purpose of searching databases and receiving records. ANSI/NISO Z39.50-1995 (ISO 23950) is one of a set of standards produced to facilitate the interconnection of computer systems. As a network protocol, the Z39.50 standard provides a set of rules that govern the formats and procedure used by computers to interact with one another. The Standard establishes the permissible sequences of events at each of the two computers. Systems specify the content and structure of information parcels that are exchanged between systems.

The standard specifies formats and procedures governing the exchange of messages between a client and a server, enabling the user to search remote databases, identify records which meet specified criteria, and to retrieve some or all of the identified records. It is concerned, in particular, with the search and retrieval of information in databases. This protocol is not used by the Internet search engines (they use http). It is more complex and more comprehensive and powerful than http.

One of the major advantages of using Z39.50 is that it enables uniform access to a large number of diverse and heterogeneous information sources. Z39.50 does offer one true interface to a variety of databases. Some of these products are starting to shift to office Unicode support, which will become increasingly important as we move into multi-lingual record displays. They are functionally rich because this kind of product can support simultaneous searching of multiple databases. This is a very valuable feature in that it greatly compresses the amount of time required to sequentially query multiple databases.

The name Z39 came from the ANSI committee on libraries, publishing and information services which was named Z39. NISO standards are numbered sequentially and Z39 is the 50th standard developed by the NISO. The current version of Z39.50 was adopted in 1995 superseding earlier versions adopted in 1992 and 1988.

12.10.7 Dublin Core or Z39.85

The Dublin Core refers to a set of metadata elements that may be assigned to web pages so as to facilitate discovery of electronic resources. Originally conceived for author-generated description of web resources at the OCLC/NCSA Metadata Workshop held at Dublin, Ohio in 1995, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organisations. The Dublin Core Workshop Series has gathered experts from the library world, the networking and digital library research communities, and a variety of content specialists in a series of invitational workshops. The building of an interdisciplinary, international consensus around a core element set is the central feature of the Dublin Core. A set of 15 core elements in Dublin Core include: Title, Creator, Subject, Keywords, Description, Publisher, Contributor, Date, Resource Type, Format, Resource Identifier, Source, Language, Relation, Coverage and Rights Management. Dublin Core has now been approved as NISO standard Z39.85 (Dublin Core Metadata).

Self Check Exercise

11) What is Z39.50 protocol? What application does it have in a library environment?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.11 ORGANISATION OF INTERNET

As mentioned above, Internet links millions of computers for communicating data from one computer to another. There are millions of host computers that communicate with each other by using standard Internet Protocols described in the last section. The data to be moved from one host to another host is broken into small pieces called packets. Each packet has a header with address of destination host. The packets of different sizes move on various networks before reaching the destination. Various packets of one file may take different routes to reach a destination. The different networks on Internet are connected with special purpose computers called routers. These routers look for destination address given on each packet and direct the packet to take best route for destination. Routers make decisions based on information that is constantly reaching them from all over the Internet. They also know from other routers about the links that are down or congested / slow, or about routers that are no longer accepting packets for certain destinations. Each packet's destination and proposed route is evaluated individually, in the blink of an eye, and sent off along the best route for that particular packet at that particular moment.

The same sort of decision-making is made for all packets that traverse the Internet. Each time a packet reaches a router, its address is examined and the packet is forwarded either to another router nearer its ultimate destination or to that destination if the router is the final router on the path. The destination computer is the one that unpacks and merges all the packets, strip away the distribution and routing addresses and pass the data.

12.11.1 Internet Addressing

Each host computer on the Internet has its own unique address. To identify a host on the Internet, three addressing systems have been evolved: A numerical system called IP addressing, a hierarchical naming system called the Domain Name System, and an addressing system called URLs, which are used for identifying sites on the web.

IP address: Each computer has a unique numerical address, such as 202.54.26.82

Domain name: Each computer must have a unique name, such as www.iitd.ac.in

Uniform Resource Locator: Address of file(s) to be accessible from a host computer

12.11.2 IP Addresses

Every host on the Internet is assigned a unique identifier called an IP address or Internet Protocol address. The IP address is a numerical address consisting of four numbers separated by periods. An IP address looks like this: 202.54.26.82 and is read as, "202 dot 54 dot 26 dot 82."

The IP address is a set of numbers that expresses the exact physical connection between a computer and the network on the Internet. They are unique and can be equated to telephone numbers in a way: a phone number uniquely describes a user's connection to the telephone network. IP addresses work somewhat similarly but are more complex than phone numbers because there are literally millions of network connections possible and because IP addresses are intended for use by computers rather than people.

An IP address consists of a 32-bit integer that's represented by four 8-bit numbers, written in base 10, separated by periods. IP addresses are organised from left to right with the left-hand octet describing the largest network organisation and the rightmost octet describing the actual network connection. IP addresses are unique on the network and allow it to know specifically which computer is to receive which electronic packet as well as from which specific computer the electronic packet came. IP address of IIT Delhi web server is 202.54.26.82. As it may be quite difficult to remember such long IP addresses, a system to translate it in domain names has been developed. The Domain Name System (DNS) serves as directory program for IP addresses and it takes care of translating IP addresses in simpler English names. The DNS for above mentioned IP address is www.iitd.ac.in.

12.11.3 Domain Names

A **domain** name locates an organisation or other entity on the Internet. For example, the domain name www.nic.in or www.hindustantimes.com

The domains names have been designed in such a way that they broadly describe organisational or geographic realities. They indicate what country the network connection is in, what kind of organisation owns it, and sometimes further details. Servers or host computers have special names for each country. All countries in the world have a country suffix, except the USA. The UAE uses .ae, New Zealand uses .nz, while Canada's is .ca.

The domain name of a host computer looks like

- Host computer name
- Organisation name
- Type of organisation
- Country name

IETF who designed the addressing system have planned a system, which looks like words. These words roughly map to a parallel system of numerical addresses called IP Addresses. Every computer on the Internet has both a domain name and an IP address, and when you use a domain name, the computers translate that name to the corresponding IP address.

The server **www.iitd.ac.in** means

- a host computer called www
- an organisation called IITD
- an academic institution (.ac stands for academic)
- located in India (.in stands for India)

Similarly, the server www.yahoo.com means a host called www, belonging to an organization called yahoo, which is a commercial organisation (.com means commercial) located in the United States (if there is no country code then it is in the United States).

In actuality, a host computer uses only numbers, turning all domain name addresses into numbers. This translation process is taken care of behind the scenes by software. The reason domain names exist in the first place is because names are more convenient for people to use and easier to remember than numbers. Domain names are used for addressing hosts rather than IP addresses.

A third level can be defined to identify a particular host server at the Internet address. In our example, 'www' is the name of the server that handles Internet requests. (A second server might be called 'www2'.) A third level of domain name is not required. For example, the fully qualified domain name could have been "totalbaseball.com" and the server assumed.

Second-level domain names must be unique on the Internet and registered with one of the ICANN-accredited registrars for the COM, NET and ORG top-level domains. Where appropriate, a top-level domain name can be geographic. (Currently, most non-U.S. domain names use a top-level domain name based on the country the server is in.) To register a U.S. geographic domain name or a domain name under a country code, contact an appropriate registrar.

More than one domain name can be mapped to the same Internet address. This allows multiple individuals, businesses, and organisations to have separate Internet identities while sharing the same Internet server.

Top-level Domain Names :- On the Internet, a Top-Level domain (TLD) identifies the most general part of the domain name in an Internet address. A TLD is either a Generic Top-Level Domain (gTLD), such as 'com' for 'commercial,' 'edu' for 'educational,' and so forth, or a Country Code Top-Level Domain (ccTLD), such as 'fr' for France or 'is' for Iceland.

A second-level domain (SLD) is the portion of a Uniform Resource Locator (URL) that identifies the specific and unique administrative owner associated with an **IP address**. The second-level domain name includes the top-level domain name. For example, in: whatis.com, 'whatis' is a second-level domain. "whatis.com" is a second-level domain name (and includes the top-level domain name of 'com'). Second-level domains can be divided into further domain levels. These sub domains sometimes represent different computer servers within different departments. More than one second-level domain name can be used for the same IP address.

The top level domain names include country names and are known as Geographic Domains and type of organisations known as Non-Geographic Domains. The geographically based top-level domains use two-letter country designations. For example, .us is used for the United States, .ca for Canada (not California), .uk for the United Kingdom or Great Britain, and .il for Israel. Each domain has a number of hosts. A few more examples are given in the following table.

Abbreviation	Meaning
au	Australia
be	Belgium
ge	Germany
jp	Japan
mx	Mexico
nz	New Zealand
uk	United kingdom

Non-Geographic Domains

There are six common top-level domain types that are non-geographical:

.com	for commercial organisations such as netcom.com, apple.com, sun.com, etc.
.net	for network organisations, such as internic.net
.gov	for parts of governments within the United States, such as nasa.gov, Oklahoma.gov, etc.
.edu	for organisations of higher education, such as harvard.edu, ucdavis.edu, mit.edu, etc.

- .mil for nonclassified military networks, such as army.mil, etc. (The classified networks are not connected to the wider Internet).
- .org for organisations that do not otherwise fit the commercial or educational designations.
- .int international organisation

The lowest level in the domain name system is the *host name*. Host names identify a computer on the Internet. For example, in the URL www.cse.iitd.ac.in, cse is the name of computer, iitd is the host name in the ac domain.

Host Name: In some instances, there are second-level domains delegated to organizations such as K-12 schools, community colleges, private schools, libraries, museums, as well as city and country governments. Examples of second-level domains are shown here:

- CC – Community colleges
- TEC – Technical colleges
- LIB – Libraries
- K12 – Kindergarten through 12th grade schools and districts
- STATE – State Government
- MUS – Museums

12.11.4 Uniform Resource Locator (URL)

A URL (Uniform Resource Locator) is the address of a file (resource) accessible on the Internet. The type of resource depends on the Internet application protocol. Using the World Wide Web's protocol, the Hypertext Transfer Protocol (**HTTP**), the resource can be an **HTML** page, an image file, a program such as a **common gateway interface** application or Java **applet**, or any other file supported by HTTP. The URL contains the name of the protocol required to access the resource, a **domain name** that identifies a specific computer on the Internet, and a hierarchical description of a file location on the computer.

On the web (which uses the Hypertext Transfer Protocol), an example of a URL is:

<http://www.iitd.ac.in/acad/library/index.html>

which describes a web page that can be accessed with an HTTP (web browser) application that is located on a computer named www.iitd.ac.in. The specific file is in the directory named /acad and subdirectory /library and is the default page in that directory (which, on this computer, is named as index.html).

An HTTP URL can be for any web page (not just a home page) or any individual file.

Self Check Exercise

12) Define IP address, Domain name and URL?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.12 INTERNET SECURITY

The security of computer and data transmitted on Internet is recognised as having a major threat. In fact, threat to security is the biggest hurdle to expansion of e-commerce on the Internet. Internet security is recognised as methods used by an organisation to protect its institutional network from intrusion. A system administrator has to ensure that intruders or hackers do not reach and manipulate data kept on the servers. The best way to keep an intruder from entering the network is to provide a security wall between the intruder and the institutional network. Most often, an intruder enters the network through a software program (such as a virus, trojan horse or a worm) or through a direct connection. Methods such as firewalls, data encryption and user authentication are used to check a hacker from entering the network.

12.12.1 User Authentication

User authentication is the first level of security mechanism to protect a computer system from unauthorised users. A user must enter a password as a digital key to enter the computer system. User authentication can be incorporated into firewall, a particular application, a document, or a network operating system such as Linux, Windows NT or Windows 2000.

Passwords are an important line of defence against intruders and it is in best interest to use this principal mode of protection at all levels. If an intruder can break into host's system using password, he or she may then be able to find other security holes in that system or use it as a means of entry into other systems on the Internet.

Password should be chosen carefully so that it cannot be guessed easily. The following principles can be considered while creating and using password:

- i) Do not use your name or a modification of your name as your password;
- ii) Do not use a word or modification of a word that occurs in a dictionary;
- iii) Do not use an acronym;
- iv) Once you have created an account password, don't share it with anyone;
- v) Change your password often, at least every three months;
- vi) Do not leave your machine unattended when you are logged in; and
- vii) Do not write your ID and password on a piece of paper or send it to your friend via e-mail.

12.12.2 Fire Walls

Firewalls are the first line of defence for an institutional network. A firewall is a combination of hardware and software that separates a local area network (LAN) into two or more parts for security purpose. It is a set of related programs, located at a network gateway server that protects the resources of a private network from users of other networks. All network connections to and from the institutional network pass through the firewall, which acts as a gatekeeper to give access to valid requests and blocks out invalid and unauthorised requests and transmissions. An enterprise with an intranet that allows its users to access the Internet installs a firewall to prevent outsiders from accessing its own private data resources and for controlling what outside resources its own users have access to.

A firewall refers to the concept of a security interface or gateway between a closed system or network and the outside Internet that blocks or manages communications in and out of the system. Basically, a firewall, working closely with a router program, examines each network packet to determine whether to forward it towards its destination. A firewall also includes or works with a proxy server that makes network requests on behalf of workstation users. The firewall is often installed in a specially designated computer separate

from the rest of the network so that no incoming request can get directly at private network resources. A number of companies make firewall products. The firewall facilitates features such as logging and reporting, automatic alarms at given thresholds of attack, and a graphical user interface for controlling the firewall. The following are the popular firewall software programs, which can be installed:

Personal computer firewall – Black ICE Agent, e-safe Desktop, McAfee Internet Guard Dog, Norton Internet Security and Zone Alarm.

Office firewalls – D-Link Residential Gateway, Linksys Etherport cable, Netgear and Sonic wall.

Corporate firewall – Check Point, CISCO Secure PIX Firewall, e-Soft-Interceptor and Sonic Wall Pro.

12.12.3 Proxy Servers

In an enterprise that uses the Internet, a proxy server is a server that acts as an intermediary between a workstation user and the Internet so that the enterprise can ensure security, administrative control and caching service. A proxy server is associated with or part of a gateway server that separates the enterprise network from outside intrusion.

A proxy server receives a request for an Internet service (such as a web page request) from a user. If it passes filtering requirements, the proxy server, assuming it is also a cache server, looks in its local cache of previously downloaded web pages. If it finds the page, it returns it to the user without forwarding it to the Internet. If the page is not in the cache, the proxy server, acting as a client on behalf of the user, uses one of its own IP addresses to request the page from the server out on the Internet. When the page is returned, the proxy server relates it to the original request and forwards it on to the user.

When a firewall is used to stop company workers from accessing the Internet, a proxy server is used to provide access. It also acts as a security device by providing buffer between inside and outside (on Internet) computers. The steps in the functioning of a typical proxy server are given below:

- i) Request for a file from a client is sent to the proxy server;
- ii) The proxy server contacts the web server to get the file;
- iii) The proxy server keeps a copy of the file in its cache; and
- iv) Forwards a copy of file to the client.

The functions of proxy server, firewall and caching can be segregated on to separate programs or combined in a single package. These programs can be hosted on to different servers or on a single server. For example, a proxy server may be in the same machine with a firewall server or it may be on a separate server and forward requests through the firewall. To the user, the proxy server is invisible; all Internet requests and returned responses appear to be directly coming to the Internet server.

12.12.4 Data Encryption

Techniques of data encryption can be used to avoid its misuse if intercepted in transit. Data encryption is a method of scrambling the data into an unreadable form before it leaves the client's machine. When the data reaches its destination, a key decodes the encrypted data into readable information. The following are the three elements associated with encryption system:

- i) A method of changing the data into code (algorithm);
- ii) A hidden place to start the algorithm (key);
- iii) Control of the key (Key management).

12.12.5 Digital Certification

Digital certificate is an electronic credit card that establishes credentials for doing business or other transaction on the net. Digital certificates are issued by a government agency called "certification authority". This certificate contains user's name, a serial number, expiration date, a copy of the certificate holder's public key and the digital signature of the certificate-issuing authority. Digital certificates are similar to watermarks on a bank note. Digital certificates not only substantiate the authenticity of a message and its sender but also alert the recipient if the message was altered while in transit.

12.12.6 Intrusion Detection System (IDS)

A specialised software termed as intrusion detection system can be deployed to monitor and detect any malicious activity on the network generated from the organisation. There are two types of intrusion detection systems currently being used, i.e., scanners and monitors. Both of these systems can be deployed onto a network or individual computers. Scanners are static IDS systems that keep a watch over a network system, like a security guard. Scanners check for things like wrong passwords, security holes and misconfigured computers. Some scanners take snapshots of the current state of the network and compare it with the snapshot taken at a later date to see if there were any changes in the system that are unwarranted. If the scanner senses such changes, it may sound an alarm or act proactively by replacing changed files with clean copies. The monitors IDS, on the other hand, are dynamic systems that look for attacks on the network while in progress. Monitor IDS are also known as threat monitors.

12.12.7 Computer Viruses

Viruses represent the single biggest threat to the computer and network security. A computer virus is a computer program that infects the computer applications or system files. When the virus becomes active, it can destroy the data on any computer. The virus does this by getting into computer's memory and from there it can copy itself to its hard disk or floppy disks.

The computer virus is a piece of programming code usually disguised as something else, which causes unexpected and usually undesirable events. A virus is often designed so that it automatically spreads to other computer users. Viruses can be transmitted as attachments to e-mail note, as downloads, or they can be present on a diskette or CD. The source of the e-mail note, downloaded file, or diskette is often the source of viruses. Some viruses wreak their effect as soon as their code is executed; others lie dormant until circumstances cause their code to be executed by the computer. Some viruses are playful in intent and effect ("Happy Birthday, Ludwig!") and some can be quite harmful, erasing data or infecting hard disk to require reformatting.

Viruses are inactive until infected applications are executed. A virus can also get activated when a computer is booted with a floppy disk that is infected by a boot sector virus. There are three main classes of virus:

File Infectors: Some file infector viruses attach themselves to program files, usually selected .COM or .EXE files. Some can infect any program for which execution is requested, including .SYS, .OVL, .PRG, and .MNU files. When the program is loaded, the virus is sent as an attachment to an e-mail note.

System or boot-record infectors: These viruses infect executable code found in certain system areas on a disk. They attach to the DOS boot sector on diskettes or the Master Boot Record on hard disks. A typical scenario is to receive a diskette from an innocent source that contains a boot disk virus. When the operating system is running, files on the diskette can be read without triggering the boot disk virus. However, if the diskette is left in the drive, and then the computer is turned-off or reload the operating system, the computer will look first in the A drive, find the diskette with its boot virus, load it, and make it temporarily impossible to use the hard disk.

Macro Viruses: These are among the most common viruses, and they tend to do the least damage. Macro viruses infect Microsoft Word application and typically insert unwanted words or phrases.

The best protection against a virus is to know the origin of each program or file that loaded into the computer or opened from the e-mail program. Since this is difficult, anti-virus software can be bought that can screen e-mail attachments and also check all the files periodically and remove any viruses that are found.

Anti-Virus Software

Anti-virus software greatly reduces chances of experiencing a viral infection in a computer. MSAV (Microsoft Anti-Virus) comes with DOS 6.22 and there are standalone products available like the Norton Anti-Virus program. Most of these programs scan all of the files currently on the hard disk to see whether they contain any viruses. They also install memory – resident programs that keep a continuous lookout for any suspicious activity that would indicate a virus.

Along with commercial products, there are a few and shareware products that are available, too. Windows and DOS users can check out the virus products offered by McAfee by either going to their FTP site at <ftp.mcafee.com> or their web site at <http://www.mcafee.com>. Another product that offers protection against viruses is called F_PROT. F-ProT is available for Windows 3.x, Windows 95/ NT, DOS, Netware and OS/2 (16-bit and 32-bit). Grisoft also offers its anti-virus programs free of charge from its site <http://www.grisoft.com>.

The virus protection software should be updated periodically for protection against new viruses that keep coming. Most anti-virus programs release their updates from their web sites.

How to Avoid Viruses

If floppies are shared with others or files are downloaded from the online services, there is no guarantee of protection against viruses. To help reduce the risk of your computer being infected, follow these tips:

- i) Run an anti-virus program and keep it updated often.
- ii) Scan floppies that are suspected to be infected.
- iii) Do not copy programs from one computer to another. Use the original distribution diskettes to install programs.
- iv) Scan your system regularly with the full scanning engine.
- v) Avoid using floppies from unknown sources.
- vi) Write-protect floppies by sliding the little tab to expose the hole on 3.5-inch disks.
- vii) Never boot computer from unknown diskettes. If it happens by accident and virus is suspected on the diskette, shut down the computer, boot up from a clean system diskette and check the system with an anti-virus program.
- viii) Utilise your anti-virus program's memory resident scanners to check all files as they are accessed, even from the Internet.

12.12.8 Virtual Private Network (VPN)

A Virtual Private Network (VPN) can be used by organisations as a means of protection from the outside attack. A VPN is a private data network that uses public networks (such as Internet), tunnelling protocols, and security procedures to tunnel data from one network to another. The data sent on a VPN is usually encrypted before it is transmitted on the net.

12.12.9 Extranet

Extranets are also deployed as a popular means to achieve Internet security. An Extranet, or extended Intranet, can be defined as a private network of linking branch offices or several cooperating organisations located outside the walls of any organisation. An Extranet service uses existing Intranet interactive infrastructure, including standard servers, email clients and web browsers. This makes Extranet far more economical than the creation and maintenance of a proprietary network. It enables trading partners, suppliers and customers with common interests to form a tight business relationship and a strong communication bond.

The Extranet can be defined as “a network that links business partners to one another over the Internet by tying together their corporate intranets”. Extranets may be used to allow inventory database searches, for example, or to transmit information on the status of an order. They are being used by businesses of all types such as banks, airlines, railways, large corporate offices having several branches, etc.

Self Check Exercise

13) What are computer viruses and anti-virus software?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

12.13 SUMMARY

The Unit introduces Internet as one of the most powerful media of communication that has completely revolutionized the modes and methods of computer and communication. It traces the history of Internet from 1957 when erstwhile Soviet Union launched its first satellite, Sputnik I, prompting US President Dwight Eisenhower to launch Defence Advanced Research Project Agency (DARPA). With the collapse of erstwhile Soviet Union in 1990, the ARPANET ceased to exist and the Internet was made available commercially to any one for asking. The Internet is defined as a network of networks that use TCP/IP protocols and that was evolved from the ARPANET. The Unit elaborates upon management of Internet through well-known organisation, i.e., Internet Activities Board (IAB) and its components that work together in a relatively well-structured and roughly democratic environment to collectively participate in the research, development and management of the Internet.

The Unit traces growth of Internet in terms of the number of hosts accessible on the Internet as a fair measure of its growth. The Internet grew from 4 hosts in 1962 to 233 million hosts in early 2004 spread over in 200 countries. The number of users of Internet has grown to 200 million. It highlights developments in Internet 2 and Internet 3 projects.

The Unit explains Internet architecture as a client-server model in contrast to the terminal-host model and peer-to-peer communications vis-à-vis the client server architecture. Client-server model and functions of client and server as pieces of interdependent software is described. Taking examples from various applications, the Unit reiterates that software

tools in client-server environment work in pairs. For every application in client-server environment, there is a client program responsible for facilitating users to interact with server programs and a server program that is responsible for hosting data and serving the client.

The Unit describes the following six methods for accessing the Internet:

- i) Dial-up access
- ii) Digital Subscribers Line (DSL)
- iii) Dedicated Leased Line
- iv) Integrated Service Digital Network (ISDN)
- v) Internet through Cable
- vi) Internet through the Mobile

It describes two types of accounts i.e., shell account v/s IP account and their advantages and disadvantages. With graphic-based applications coming in a big way, shell accounts are not being preferred.

The Unit describes Internet Service Providers (ISP) as companies that provide access to the Internet and the Internet Service Providers (ISPs) in India. There are around 200 ISPs that are providing services to 2.5 million users in the country. Important ISPs in India are enlisted. The Unit provides hardware and software requirements both at the server end as well as at the client's end.

The protocols are described as a formal description of formats and rules that two or more computers must follow to exchange data. The Unit elaborates upon the number of protocols including: TCP / IP, HTTP, FTP, SLIP, PPP, X39.50 and Z39.85.

It describes IP address, domain name and Universal Resource Locator (URL) as addressing systems used for the Internet. Lastly, the Unit describes security of computers over the Internet. It elaborates a number of measures to secure computers over the Internet, namely, use of: user authentication, fire walls, proxy servers, data encryption, digital certification, Intrusion Detection Systems (IDS), anti-virus software, Virtual Private Networks (VPN) and Extranets.

12.14 ANSWERS TO SELF CHECK EXERCISES

- 1) DARPA and NSF played an important role in the growth and development of Internet especially for infrastructure development and for supporting various networking-related research programmes. NSF provided funds for developing a wide area network infrastructure to support the general academic and research community. DARPA's ARPANET was initially made accessible to universities and research institutions to facilitate easier and faster communication among scientists and researchers. With winding of operations of ARPANET in 1990, the combined infrastructure of ARPANET and NSFNET became available commercially.
- 2) The Internet does not have a central authority to control its activities. There are, however, several well-known organizations that work together in a relatively well-structured and roughly democratic environment to collectively participate in the research, development, and management of the Internet. The Internet Activities Board (IAB) was created in 1983 to guide the evolution of the TCP/IP Protocol suite and to provide research advice to the Internet community. IAB has two primary components: the Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF). While the IETF is primarily responsible for further evolution of the TCP/IP protocol suite, the Internet Research Task Force (IRTF) continues to

organize and explore advanced concepts in networking under the guidance of the Internet Activities Board.

The Internet Assigned Numbers Authority (IANA) and Internet Registry (IR) facilitate central allocation of network and autonomous system identifiers. The Internet Registry also provides central maintenance of the Domain Name System (DNS) root database which points to subsidiary distributed DNS servers replicated throughout the Internet. Besides, there are a number of Network Information Centers (NICs) located throughout the Internet to serve its users with documentation, guidance, advice and assistance.

- 3) The term Internet has been coined from two terms, i.e., interconnection and network. A network is simply a group of computers that are connected together for sharing information and resources. Several such networks have been joined together across the world to form what is called as the Internet. The Internet is thus a network of networks. It refers to the vast collection of interconnected networks that use the TCP/IP protocols and that evolved from the ARPANET of the late 1960's and early 1970's. Use of the common protocol called TCP / IP makes it possible for different platforms and operating systems to be a part of the Internet.
- 4) Client-server model distributes the processing of a computer application between two computers, the client and the server. The client is normally a PC. The application program accesses data and performs processing on the server and using the data obtained via the server more processing tasks are performed at the client. In a host-terminal model, a server computer does all the work and terminals give each user access to the contents on the server. The advantage of a host-terminal model is that all the maintenance is performed at one place, i.e., on the server. As computers became more powerful and readily available, the client-server model became popular.
- 5) Packet switching is a switching system that uses a physical communication connection only for the period of transmission of data. Data are disassembled into packets and reassembled at the receiving end of the communications link; packets may travel over many diverse communications links to get to the common end point. Circuit switching, in contrast, is a type of communication in which a dedicated channel (or circuit) is established for the duration of a transmission. The telephone system is a circuit-switching network because it consists of wire segments linked together to create a single unbroken line for each telephone call.
- 6) One of the six methods mentioned here can be used for accessing the Internet: i) Dial-up access; ii) Digital Subscribers Line (DSL); iii) Dedicated Leased Line; iv) Integrated Service Digital Network (ISDN); v) Internet through Cable; and vi) Internet through the Mobile. Dial-up access is most common way to access the Internet.
- 7) A Digital Subscribers Line (DSL) connection works better when it is closer to the Internet provider's central office. The connection is faster for receiving data than it is for sending data over the Internet. The DSL is available only in metropolitan cities.
- 8) A protocol is a standardized means of communication among machines across a network. Protocols allow data to be taken apart for faster transmission, transmitted, and then reassembled at the destination in the correct order. The protocol used determines the way errors are checked, the type of compression, the way the sender indicates the end of the transmission, and the way the receiver indicates that the message has been received. Protocols can describe low-level details of machine-to-machine interfaces (e.g., the order in which bits and bytes are sent across a wire) or high-level exchanges between application programs, e.g., the way in which two programs transfer a file across the Internet. Standard protocols makes the Internet work.
- 9) TCP/IP (Transmission Control Protocol / Internet Protocol) This is the suite of protocols that defines the Internet. TCP/IP is a standard format for transmitting data from one

- DNS** : Short for Domain Name Server, used to map names to IP addresses and vice versa. Domain Name Servers maintain central lists of domain name / IP addresses.
- DSL** : Short for Digital Subscriber Line. A method for moving data over regular phone lines. A DSL circuit is much faster than a regular phone connection, and the wires coming into the subscriber's premises are the same (copper) wires used for regular phone service.
- Data Encryption** : It is a security procedure that encodes data so that it cannot easily be understood. To be usable, data must be decrypted into its original form by reversing the procedure that was used to encrypt it.
- Dublin Core** : A metadata element set intended to facilitate discovery of electronic resources. Dublin Core is intended to be usable by both non-cataloguers and specialists alike, to provide an economic alternative to more elaborate descriptive models such as full MARC cataloguing. The Dublin Core can be mapped to the MARC record and a variety of output structures can be generated.
- Fire Wall** : A combination of hardware and software that separates a LAN into two or more parts for security purposes. Fire walls are often used by institutions which house their web site on a server located in the institute. The Fire wall provides a measure of security that those accessing the company's web site will not be able to access private data that is housed on the same network.
- IDS** : Stands for Intrusion Detection System, designed to protect a specific portal, volume or area, using technologies designed to sense movement, sound or a specific act such as opening a door. This security alarm system consists of various types of sensors (vibration, capacitance, volumetric, etc.) to detect the unauthorized intrusion into a facility. Typical systems include ultrasonic, infrared, microwave sensors, and door switches. IDS systems can be local or connected to a central station.
- ISDN** : Integrated Services Digital Network is a system of digital phone connections that allows voice and data to be transmitted simultaneously across the world using end-to-end digital connectivity. There are two basic types of ISDN service: Basic Rate Interface (BRI) and Primary Rate Interface (PRI). BRI is a basic service and is intended to meet the needs of most individual users. PRI is intended for users with greater capacity requirements.
- ISP** : Internet Service Provider, a company that provides access to the Internet. Typically these service providers also offer web hosting services to their users.
- Internet** : A global network of networks connecting millions of computers using TCP/IP protocol suite to support a number of applications for information sharing and retrieval including the World Wide Web. Each Internet computer, called a host, is independent. Its operators can choose which Internet services to provide to its local users and which local services to make available to the global Internet community. Unlike online services, which are centrally controlled, the Internet is decentralised by design.
- Modem** : Modulator / demodulator, a device that allows computers to communicate over telephone lines by converting digital signals to

analogue, and vice versa. Modems can be either external, outside your computer, or internal, inside your computer.

OSI : Stands for Open Systems Interconnection. OSI is a seven-layer architecture model for communications systems developed by the ISO for the interconnection of data communications systems. Each layer uses and builds on the services provided by those below it.

PSTN : Short for Public Switched Telephone Network, this is the standard telephone service that most homes use. It is also referred to as POTS, or Plain Old Telephone Service.

Packet Switching : A data transmission method whereby data is transmitted in packets through a network to a remote location. The packet switch sends packets from different data conversations along the best route available in any order. At the other end, the packets are reassembled to form the original message which is then sent to the receiving computer. Because packets need not be sent in a particular order, and they go through any route as long as they reach their destination, packet switching networks can choose the most efficient route and send the most efficient number of packets down that route before switching to another route.

Prodigy : Prodigy is one of the providers of online services in USA. In addition to providing access to the Internet and World Wide Web, it offers news, weather, sports, computing support, popular publications, email, financial resources, as well as other resources.

Proxy server : A proxy server is a server that acts as an intermediary between a computer and the Internet. It ensures security, administrative control, and caching. Caching speeds up Internet access. If one or more Internet sites are frequently requested, they are kept in the proxy's cache, so that when a user requests them, they are delivered directly from the proxy's cache instead of from the original Internet site.

SLIP/PPP : Serial Line Internet Protocol / Point-to-Point Protocol. Allows a computer with a modem to connect to the Internet using TCP/IP. Commonly used to connect to an Internet Service Provider.

SMTP : Stands for Simple Mail Transport Protocol. The main protocol used to send electronic mail on the Internet. SMTP consists of a set of rules for how a program sending mail and a program receiving mail should interact.

TCP/IP : Stands for Transmission Control Protocol / Internet Protocol, is a suite of protocols that defines the Internet. Originally designed for the UNIX operating system, TCP/IP software is now available for every major kind of computer operating system.

USB : It is the abbreviation of Universal Serial Bus. USB is a standard port that enables a user to connect external devices (such as digital cameras, scanners, and mouse). The USB standard supports data transfer rates of 12Mbps (million bits per second), a vast improvement over the serial port standard. Aside from speed advantages, USB devices can be connected or disconnected without the need to restart the computer.

VPN : A VPN, or Virtual Private Network, is a network that is constructed by using public wires to connect nodes, set up solely for the users of a single company or organization. These networks use encryption

and other security mechanisms to ensure that only authorized users can access the network and that data cannot be intercepted.

WAIS

: Stands for Wide Area Information Server, WAIS is a commercial software package that allows indexing of huge quantities of information and then making their indices searchable across networks such as the Internet. A prominent feature of WAIS is that the search results are ranked (scored) according to how relevant the hits are to the search topic.

Z39.50

: Z39.50 is a national standard defining a protocol for computer-to-computer information retrieval. Z39.50 makes it possible for a user of one system to search and retrieve information from other Z39.50 computer systems without knowing the search syntax used by other systems. Z39.50 is an American National Standard that was originally approved by the National Information Standards Organization (NISO) in 1988.

12.16 REFERENCES AND FURTHER READING

ComputerScope. *The Worlds Leading Resource for Internet Trends and Statistics*. (http://www.nua.ie/surveys/how_many_online/world.html)

Dawson, A. (1997). *The Internet for Library and Information Professionals*. London: Library Association Publishing.

Dern, Daniel (1994). *The Internet Guide for New Users*. New York: McGraw Hill.

Ellsworth, Jill and Barron, Billy, [et al.] (1997). *The Internet 1997*. Indianapolis: Sams.net Publishing.

Hahn, Harley (1997). *Internet: Complete Reference*. 2nd ed. New Delhi: Tata McGraw Hill.

Internet Economy Indicator. (<http://www.internetindicators.com/factfigure.html>).

Internet World Stats: Usage and Population Statistics. (<http://www.InternetWorldStats.com>)

Issues of Growth and Evolution in the Internet. (<http://www.zvon.org/tmRFC/RFC1380/Output/chapter2.html>)

Johnson, Dave (1998). *Internet Explorer 4: Browsing and Beyond*. New Delhi: Tata McGraw Hill.

Kane, Kevin. *Choosing your ISP and Internet Connection Type*. (http://www.arts.uwaterloo.ca/ACO/newsletters/s01/articles/isp_and_connection_type.html).

Kumar, PSG and Vashishth, CP. CALIBER – 99: Academic libraries in Internet era: Paper presented at the *Sixth National Convention for Automation of Libraries in Education and Research*, Nagpur, 18 – 20 Feb., 1999. Ahmedabad, INFLIBNET, 1999.

Living Internet (<http://www.livinginternet.com/>).

McBride, P.K. (1999). *Internet Made Simple*. 2nd ed. Oxford: Butterworth-Heineman.

Mehta, Subhash. (1996). *Understanding and Using Internet*. Delhi: Global Business Press.

Nair, R. Raman.(2002). *Accessing Information through Internet*. New Delhi: Ess Ess Publications.

Parekh, Harsha (1999). *Internet in the Scholarly Communication Process*. Mumbai: Mukherjee Knowledgeware Association.

Randall, Neil. (2002). *Teach Yourself the Internet in a Week*. New Delhi: Prentice Hall of India.

Turpen, Aaron. *Different Internet Connection Types and their Pros and Cons*. Teachnology, Inc. (<http://www.teach-nology.com/tutorials/connections/>)

Types of Internet Connection. (<http://www.dmine.com/bbscorner/telcon.htm>). Last updated on March 5, 2004

Whittaker, Jason (2002). *Internet: Basics*. London: Routage.



ignou
THE PEOPLE'S
UNIVERSITY

UNIT 13 SEARCH ENGINES

Structure

- 13.0 Objectives
- 13.1 Introduction
- 13.2 Search Engines: Definitions
- 13.3 Search Engines: Evolution
- 13.4 How Do Search Engines Work?
 - 13.4.1 The Robot or Spider
 - 13.4.2 The Database
 - 13.4.3 The User Interface or the Agent
- 13.5 Search Engines: Categories
 - 13.5.1 Primary Search Engines
 - 13.5.2 Meta Search Engines
 - 13.5.3 Subject or Web Directories
 - 13.5.4 Hybrid Search Engines
 - 13.5.5 Subject Gateways or Subject Portals
- 13.6 Choosing a Search Engine
 - 13.6.1 Ease of Use
 - 13.6.2 Comprehensiveness
 - 13.6.3 Quality of Contents
 - 13.6.4 Control Over the Search
 - 13.6.5 Flexibility in Searching
 - 13.6.6 Assessment of Relevance
 - 13.6.7 Informative Presentation of Results
- 13.7 Searching the Web: Search Techniques
 - 13.7.1 Searching Subject Directories
 - 13.7.2 Searching Search Engines
- 13.8 Search Results
 - 13.8.1 Result Lists
 - 13.8.2 Result Descriptions
 - 13.8.3 Relevance Ranking
- 13.9 Meta Tags
- 13.10 Search Engines: Evaluation
 - 13.10.1 Database of Web Documents
 - 13.10.2 Capabilities of a Search Engine
 - 13.10.3 Results Display
- 13.11 Important Search Engines
 - 13.11.1 Primary Search Engines
 - 13.11.2 Web Directories
 - 13.11.3 Meta Search Engines
 - 13.11.4 Subject Portals or Subject Gateways

ignou
THE PEOPLE'S
UNIVERSITY

- 13.12 Summary
- 13.13 Answers to Self Check Exercises
- 13.14 Keywords
- 13.15 References and Further Reading

13.0 OBJECTIVES

After reading this Unit, you will be able to know the following components of Internet search engines:

- search engines and their evolution;
- how search engines work;
- components of a search engine;
- categories of search engines;
- search techniques;
- meta data and search engines;
- evaluation of search engines; and
- important search engines.

13.1 INTRODUCTION

The growth of the Internet has led to a paradoxical situation. While on the one hand there is a colossal amount of information available on the Internet, on the other hand sheer volume of unorganised information makes it difficult for the users to find relevant and accurate information in a speedy and efficient manner. Internet can be said to be the most exhaustive, important and useful source of information on almost all aspects of knowledge hosted on millions of servers connected to Internet around the world. It is a known fact that there are neither defined policies for hosting information nor is there a centralised database for organising and searching the information available on the Internet. This makes the Internet as the most diverse and unorganised source of information. Searching for specific information is the main purpose of using Internet for several users. However, with availability of excessive information, it has become very difficult for a common user to search for precise and relevant information on the Internet. To tackle this situation, computer scientists came up with search tools that search through the information on the Internet to churn out required information by a user. There are varieties of search, resource discovery and browsing tools that have been developed to support more efficient information retrieval. Search engines are one of such discovery tools.

Search engines use automated programs, variably called bots, robots, spiders, crawlers, wanderers and worms developed to search the web. The robots traverse the web in order to index websites. Some of them index websites by title, some by Uniform Resource Locators (URLs), some by words in each document in a website, and some by combinations of these. These search engines function in different ways and search different parts of the Internet.

13.2 SEARCH ENGINES: DEFINITIONS

Search Engine is a generic term used for the software that 'searches' the web for pages relating to a specific query. 'Google' and 'Excite' are two examples of common search

engines that index and search a significant part of the web. Several websites have their own search engines to index their own websites. The World Wide Web has several sites dedicated to indexing of information on all other sites. These sites allow a user to search the web for any word or combination of words for information resources on the web.

A search engine is a computer program that searches documents on the Internet containing terms being searched by a user. A search engine can be defined as a tool for finding, classifying and storing information on various websites on the Internet. It can help in locating information of relevance on a particular subject by using various search methods. It is a service that indexes, organises, and often rates and reviews websites. It helps users to find the proverbial needle in the Internet haystack. Different search engines work in different ways. Some rely on people to maintain a catalogue of websites or web pages; others use software to identify key information on sites across the Internet. Some combine both types of service. Searching Internet with different search engines for the same topic, therefore, provides different results. Fig.13.1 shows number of hits for 25 single word queries conducted on nine search engines. Google found more number of hits than any other search engine.

Total Hits from 25 Searches

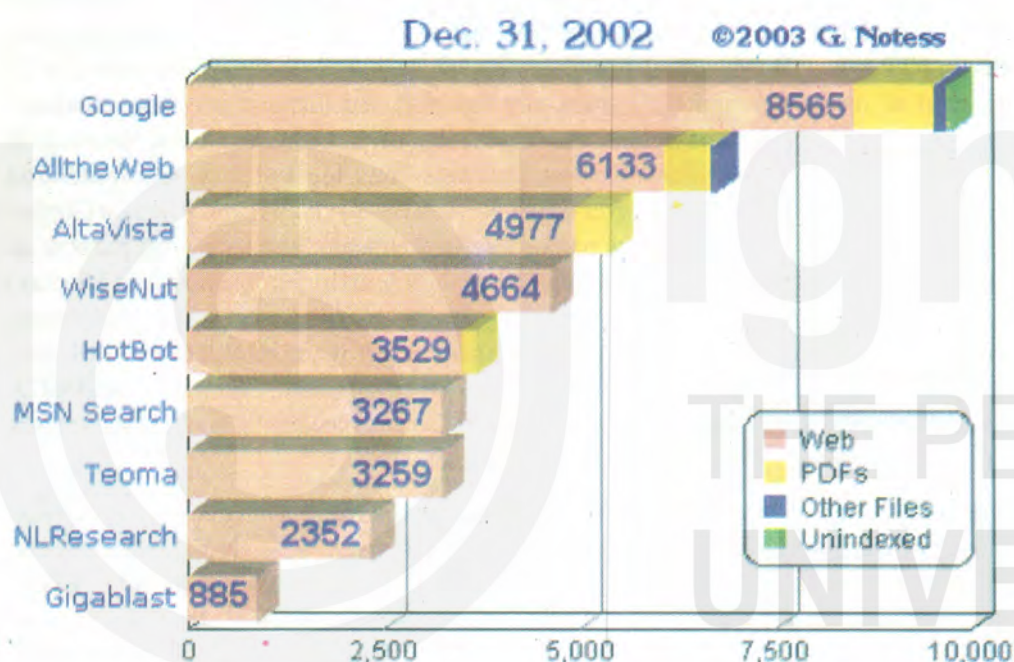


Fig. 13.1: Number of Hits for 25 Single Word Queries Conducted on Nine Search Engines.

(Source: <http://searchengineshowdown.com/stats/size.shtml>)

Search engines are also defined as online utilities that quickly search thousands of web documents for a word or phrase being searched. Although there are some subscription-based search engines, most of them operate on profits from advertisements. It should be noted that no single search engine has the contents of every web page on the Internet. Instead, each search engine defines its scope in terms of contents for web pages that it would host. Moreover, some search engines index every word on every page. Others index only a part of the document. Full-text search engines generally pick up every word in the text except commonly occurring stop words such as 'a', 'an', 'the', 'is', 'and', 'or' and 'www'. Some of the search engines discriminate upper case from lower case; others store all words without reference to capitalisation. A user, therefore, gets different results from different search engines because of these reasons.

Search engines are usually accessed using web clients called web browsers. Each search engine provides different search options and has its own peculiarities. Search engines also differ greatly in the types of resources they allow a user to search. Many search engines offer both search and browse interfaces.

Self Check Exercise

1) Define search engines. Are all search engines alike?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

13.3 SEARCH ENGINES: EVOLUTION

The Archie, developed in 1990 by Alan Emtage, a student at McGill University in Montreal, can be considered as the first search engine that was used for indexing and searching files on FTP server. It was actually an indexing spider that visited each anonymous FTP site, read all directories and file names, and then indexed them in one large database. Inspired with the success of Archie, the University of Nevada developed Veronica in 1993. VERONICA (Very Easy Rodent-Oriented Netwide Index to Computerized Archives) was developed at the University of Nevada to search all menu items on Gopher servers. Soon, another user interface named Jughead appeared with the same purpose as Veronica. The JugHead (Jonzy's Universal Gopher Hierarchy Excavation And Display) was a powerful Gopher search tool written by Rhett 'Jonzy' Jones. It was a computer program that searched a specified Gopher site (not all). It searched directory titles only, not the text of resources that appeared on the Gopher submenus. Archie, VERONICA and Jughead have now disappeared, but before the web's spectacular growth, these tools were real workhorses for searchers on the Internet.

Soon after launch of World Wide Web in 1993, the first robot, called World Wide Web Wanderer, was introduced by Matthew Gray to search the web. In October 1993, Artijn Koster developed an Archie-like Indexing tool for the web, called ALIWEB. It did not use a robot to collect the metadata, instead, it allowed users to submit the websites they wanted to be indexed by ALIWEB with their own descriptions and keywords. By December of 1993, three full-fledged robot-fed search engines had surfaced on the web, i.e.; JumpStation, the World Wide Web Worm, and the Repository-Based Software Engineering (RBSE) spider. JumpStation gathered information about the title and header from web pages and retrieved them using a simple linear search. As the web grew, JumpStation slowed to a stop. The WWW Worm indexed titles and URLs. The JumpStation and the World Wide Web Worm did not use any ranking method to list their search results, results were listed in the order they were found. The RSBE spider did implement a ranking system.

The 'Excite' was a by-product of the project called Architext that was started in 1993 by six Stanford undergraduates. They used statistical analysis of word relationships to make searching more efficient. The Excite search software was released by mid-1993. However, the technique used by the Excite seems irrelevant because the spiders were not intelligent enough to understand what all the links meant. The EInet Galaxy Web Directory was launched in January, 1994. The EInet Galaxy became a success since it also contained Gopher and Telnet search features in addition to its web search feature.

In April 1994, David Filo and Jerry Yang created Yahoo as a collection of their favourite web pages. As their number of links grew, they had to reorganise and develop a searchable directory. The Yahoo directory provided description with each

URL as an improvement to the Wanderer. Brian Pinkerton of the University of Washington launched the WebCrawler on April 20, 1994. It was the first crawler that indexed entire pages. In 1997, Excite bought out WebCrawler, and AOL began using Excite to power its NetFind. WebCrawler opened the door for many other services to follow the suit.

Three important search engines, namely Lycos, Infoseek and OpenText appeared soon after Web Crawler was launched. Lycos was the next major search engine developed at Carnegie Mellon University in July 1994. On July 20, 1994, Lycos was launched with a catalogue of 54,000 documents. By August 1994, Lycos had identified 394,000 documents and by November 1996, Lycos had indexed over 60 million documents, more than any other web search engine. In October 1994, Lycos ranked first on Netscape's list of search engines by finding the most hits on the word 'surf'. Infoseek was also launched in 1994. In December 1995, Netscape started using Infoseek as its default search engine. AltaVista was also launched in December 1995. It brought many important features to the web searching. They were the first to allow natural language queries and advanced searching techniques.

The LookSmart directory commenced functioning in 1996. The Inktomi Corporation came about in May 1996 with its search engine called Hotbot. It was bought by Yahoo. Ask Jeeves was launched in April 1997 followed by the Northern Light.

1998 witnessed the launch of Google, the most powerful search engine till date. The Google ranks its pages based on number of inbound links to a page. Google has become so popular that major portals such as AOL and Yahoo have used Google to search their directories. In 1998, three major search engines and directories were launched, they were: MSN search, Open Directory and Direct Hit.

Disney released the Go Network in 1999. Fast released its search technology in the same year, and was considered the closest competitor to Google. In 2000, the Teoma search engine was released, which uses clustering to organise sites by subject-specific popularity. In 2001, Ask Jeeves bought Teoma to replace the Direct Hit search engine.

LookSmart bought the WiseNut search engine in 2002 to power their new search product. In 2003, Google began to introduce semantic elements into its search product thereby bringing improvements in its search result. Overture purchased AllTheWeb and AltaVista. Yahoo bought Inktomi and Overture. In 2004, MSN dropped LookSmart in favour of Inktomi and Yahoo dumped Google in favour of its own search engine. Yahoo! has built a new database separate from the Inktomi database, that replaced both AltaVista and AllTheWeb in March 2004.

The brief history given here does not include many smaller search engines that came, marginalised and disappeared.

Self Check Exercise

- 2) Justify the statement that "Archie, VERONICA and Jughead were the real workhorses for searchers on the Internet before advent of the Web".
- 3) Which search engine can be considered as the first search engine for the World Wide Web?

Note : i) Write your answers in the space given below.

- ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

13.4 HOW DO SEARCH ENGINES WORK?

Search engines do not really search the World Wide Web directly. Instead, they search their own databases consisting of the keywords or full text of web pages that were earlier selected and picked-up from billions of web pages residing on servers all over the world. When a user searches the web using a search engine, it always searches an old copy of the real web page that is residing on the server of a search engine. When a user clicks on the links provided in a search engine's search results, he / she is directed to the current version of the page. A typical search engine has the following three components:

13.4.1 The Robot or Spider

Variably known as bot, robot, spider, crawler, web wanderer or indexer, they are automatic computer programs that traverse the World Wide Web information space. These programs move from one web page to another by visiting links embedded on each web page it finds and in the process builds an index to visited web pages. This process can be compared to citation searching, where a user follows a reference within a journal article to another article on the same topic. The spider is resident on a host computer and uses the HTTP protocol (like any web surfer) to connect to remote web pages. Spiders may be configured to retrieve files in all formats. They crawl on the web to discover new resources, to index the web for keywords and, to look for dead links for automated maintenance. Unlike a virus, a robot does not physically moves from computer to computer; it simply visits sites, like a user, and requests for documents to be indexed. Main functions of a robot or a spider are indexing of web pages, HTML validation, link validation, identifying new information and mirroring websites.

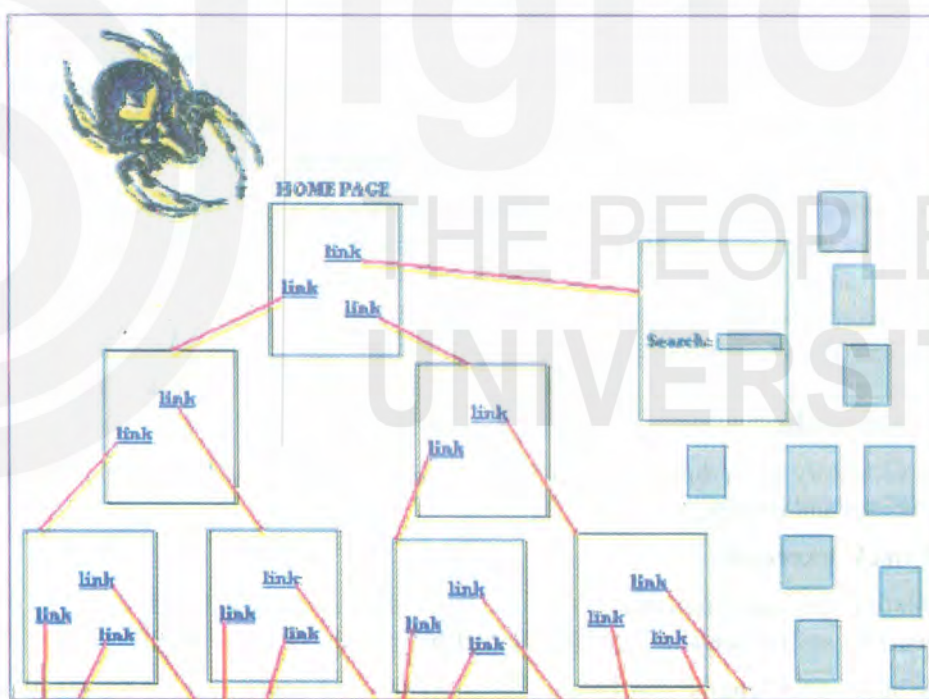


Fig. 13.2: Thematic Presentation of Functioning of Web Robots

Different robots use different strategies to index the web. In general, most search engines start from a meta resource site or a subject portal that contain links to several other resources. They scan the web constantly, keeping track of new documents that appear afresh and deleting documents that are duplicate. Most search engines permit users to submit URLs manually that are queued and visited by the robot later. While some search engines are highly comprehensive in their coverage and index all kinds of resources available on the web pages, news groups, discussion forum, Gopher, listservs and FTP sites, etc., others restrict themselves to indexing web pages only. Similarly, some search engines are programmed to traverse the net daily, others are programmed to visit predefined sites at predefined intervals.

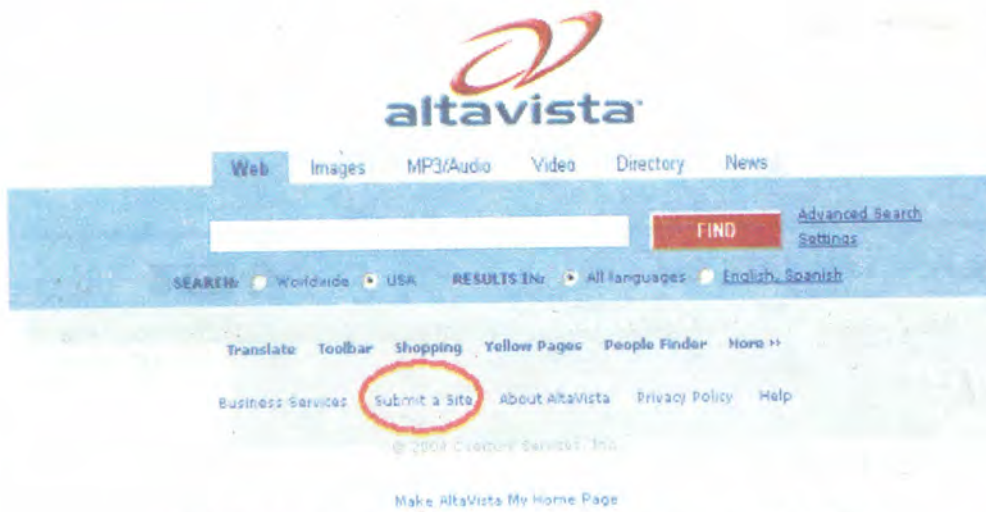


Fig.13.3: Search Interface for Alta Vista with “Submit a Site” Link

Search engines use their own spider software that are designed to harvest indexing information from websites on specified criteria. Similarly, search interfaces are customised to highlight special features of various search engines. Each uses its own algorithm or set of rules that determines how the search engine selects, indexes, and retrieves sites. Some types of pages and links (such as pornographic sites) are excluded from most search engines by policy. Besides, there are sites that do not allow search engines to index them. Moreover, search engines cannot index database-driven sites. Such sites are referred to as the ‘Invisible Web’ or ‘Hidden Web’, information that is priced and hidden behind databases.

After spiders find pages from web, they pass them on to another computer program for ‘indexing’. This program identifies the text, links and other content in the page and stores it in the search engine database’s files.

Self Check Exercise

- 4) What are spiders? Do all spiders function in the same way?
- 5) Name various components of a search engine.

Note : i) Write your answers in the space given below.
 ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

13.4.2 The Database

A robot or a spider harvests indexing information from web pages that it visited in its database or catalogue that lists URLs, titles, headers, words from title and text, first lines, abstracts, and some times even full-text. The robot performs search in a way that the most popular sites are found and indexed first. The resulting database, that stores millions of web pages, forms the index that is searched by the users. The size of this database determines the comprehensiveness of a search engine. Most search tools also create a separate database containing records consisting of a web page, URL, title and a summary. When a user retrieves results from a search tool, a summary record is displayed for the users.

Search engines either update their databases cumulatively or rebuild it completely after a definite interval. Lycos, for example, updates information on new and existing URLs cumulatively rather than rebuilding it. Excite, on the other hand, rebuilds its database periodically by sending its spider to bring back the contents of pages identified from the 'new sites' every week.

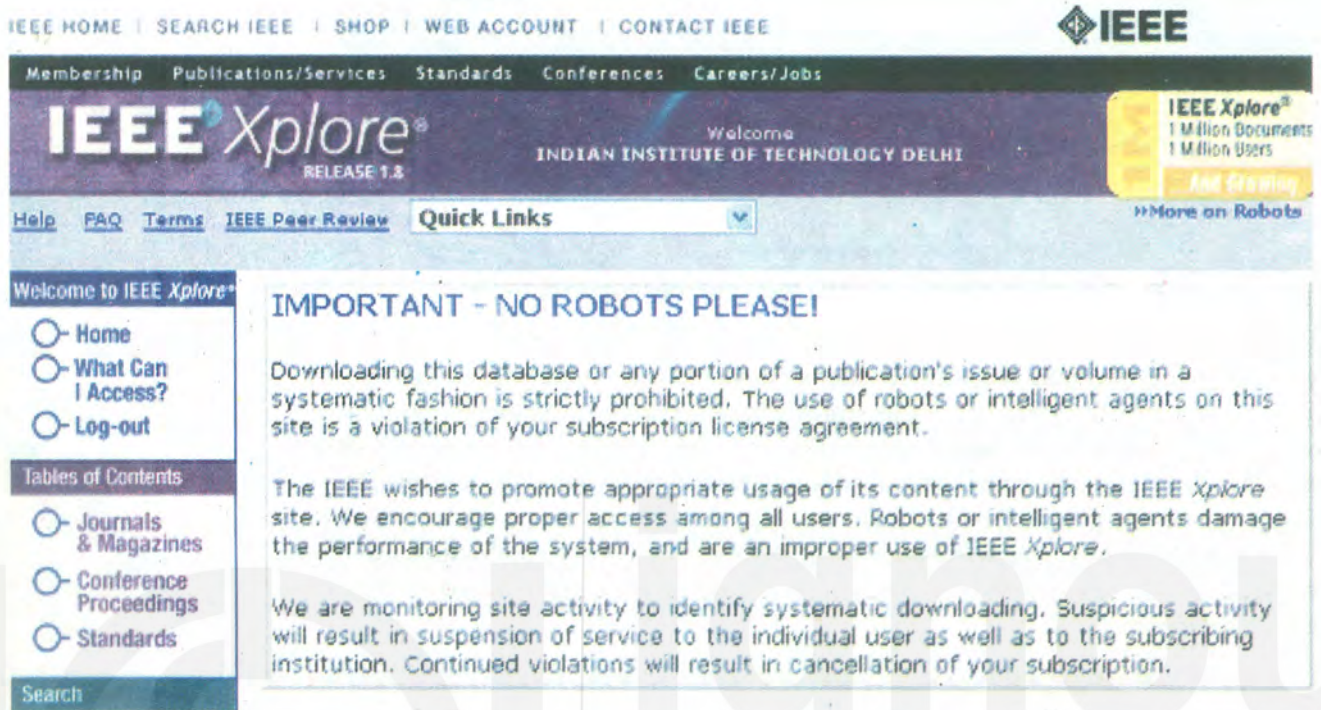


Fig. 13.4: IEEE Xplore with Instructions to Robots or Intelligent Agents Prohibiting Systematic Download of Contents from the Site

13.4.3 The User Interface or the Agent

The user interface or agent is a software program that accepts queries from users and searches them through the database consisting of index of millions of pages. The agent matches the query with the database, finds hits and ranks them in order of relevance. The results consisting of web links and brief descriptions are arranged in order of relevance and are presented to the users. Amongst items with same relevance, the most popular sites are listed first.

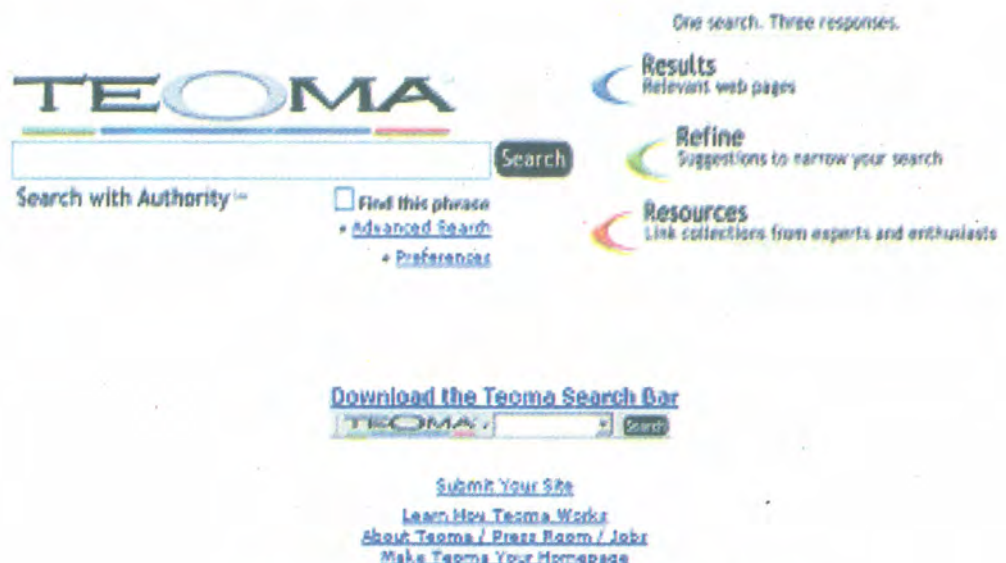


Fig. 13.5: Search Interface for TEOMA Search Engine

13.5 SEARCH ENGINES: CATEGORIES

Most search engines facilitate several functionalities besides being a tool for finding websites. They provide 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 business and individuals. Search engines can be divided into the following categories:

13.5.1 Primary Search Engines

Primary search engines deploy computer programs called web crawlers or spiders, web wanderers and web worms, to traverse the web and scan websites for words, phrases, or the whole site so as to generate a database of web pages. Search engines do not actually search the web in real time, but uses database of web page collected by its robots hypertext links that are updated on a regular basis. Primary search engines are the most commonly used search engines. These vary to a great extent in terms of:

- database size: width and depth of websites indexed by their 'spiders'.
- database content: Full-text or metadata, i.e., URL, keywords, title, description, etc.
- syntax used: word search, Boolean search, phrase search and other advanced features.
- ranking of results: paid sites, recent update, popularity, etc.

Primary search engines can also be divided into the following four categories according to the method their robots use for collecting information for their index databases:

Automated Robots: The automated robots scan a large part of web wherever they are allowed.

Designated Robots: The designated robots (like those used by ALIWEB or W3 catalogue) are programmed to scan only specific sites on the web rather than traverse the entire web. Sites using 'designated robots' provides users to submit their websites to the search engine. On submission of a URL, the new URL gets added to the robot's queue of websites to be visited on its next foray out onto the web. Even if a user does not submit its site, a designated robot may pick it up from sites that provide a link to other sites.

Breadth-oriented Search Engines: Some robots concentrate more on top-level resources that tend to store larger subject-oriented index engines such as JumpStation II. A query conducted on "breadth-oriented search engines" would result in a fewer retrievals with a higher percentage of those being meta resources or subject portals.

Depth-oriented Search Engines: Depth-oriented robots (like web crawler) follow links to deeper levels. They pull out individual items located on a server's indexes and follow links to other servers. Depth-oriented robots have the tendency to retrieve duplicates or false hits also. Depth-oriented search engines have a tendency to catalogue too much information.

Most search engines prompt a user to enter one or more search terms, press 'Enter' or click at 'search' or 'go' to begin a search. Most primary search engine offer three basic search options, i.e., any word, all words and exact phrase. Searching '**any word**' results in a list of all the indexed pages that contain any of the search terms. For example, a search for "Mahatma Gandhi birthdate" would return all the pages containing the word 'Mahatma' as well as all the pages containing the word 'Gandhi' and 'birthdate'. This type of search yields a large number of search results. 'Any words' search can be useful when a user is unsure of the search terms he should use for his search. '**All words**' search lists all indexed pages that contain all the terms being searched, in any order. Once again, there is a strong possibility of off-topic results. '**Exact phrase**' search lists all indexed pages that contain the exact phrase entered, ignoring punctuations. Sites on the same topic that are described using a different phrase will not appear. However, even this type of search can yield false results since the exact phrase may have been used in different context.

Most search sites offer more than these three simple search types, either on their home pages or on a page named 'Power Search' or 'Advanced Search'. Different sites support different features. The results of a search may differ to a great extent depending on the option chosen.

Search results are usually ranked in order of relevance, i.e., the number of times the search terms appear in a document, location of search term in a page (title, description, etc.), the number of inbound links and how closely the document appears to match the concept being searched.

13.5.2 Meta Search Engines

Meta search engines, also known as mega indexes, do not have their own databases. Instead, they have access to other primary search engines. Meta search engines may conduct a search on a single primary search engine or it may search a number of primary search engines simultaneously and present the unified results as a single package. Such search engines are also called multi-threaded search engines. Ask Jeeves, MetaCrawler, Savvy Search, @Once!, All-in-One Search Page, Galaxy, Internet Sleuth, Magellan, Net Search, Dogpile, Metacrawler, Metafind, Metasearch, ixquick.com are some of the better-known meta search engines.

A meta search engine is defined as a search engine that searches the databases of several other search engines at the same time to locate web pages that match search terms given by a user. Unlike the primary search engines and directories, meta-search engines do not have their own databases, i.e., these search engines do not collect web pages, do not accept URL additions, and do not classify or review websites. Instead, they send queries simultaneously to multiple search engines and / or web directories. Many of the meta-search engines integrate search results: duplicate findings are merged into one entry, some of them rank the results according to various criteria, some allow selection on search engines to be searched.

Before conducting a search on a meta-search engine, it is important to find out which search engines are included by the meta search engine. Most meta search engines act as default search engines to majority of popular search engines, such as Google, AltaVista, Excite, Lycos and GO.com.

Successful use of a meta search engine depends on the current status of each of the primary search engines being used. Some primary search engines may be too busy at that time and some may be unreachable. It may be noted that a query submitted to a meta-search engine, with its uniform search interface and syntax, is to be applied against the diversity of individual search engines. It is, therefore, impossible for one meta search engines to take advantage of all the features of the individual search engines. Boolean searches, for example, may produce varied results. Phrase searches may not be supported. Other features, such as query refinement, are sacrificed in a meta search engine.

Moreover, meta-search engines generally do not conduct exhaustive searches and do not bring back all the pages from each of the individual search engines. They only make use of the top 10 to 100 hits from each of them. While this is sufficient for most searches, individual search engines must be consulted if one needs to go beyond the top hits as determined by the meta-search engines. Some meta-search engines facilitate this by providing query links back to the individual search engines.

Specialised search engines are primary search engines that focus on a small or specialised segment of the Internet. Examples of specialised search engines are Direct Search (<http://www.freepint.com/gary/direct.htm>), Beaucoup (<http://www.beaucoup.com/>), Hoovers Online, (<http://www.hoovers.com/>) and Sirius (<http://www.sirus.com/>).

13.5.3 Subject or Web Directories

Directories are the Yellow Pages of the Internet. They contain information that has been submitted to them by their indexers or by users who submit entries. The subject directories are often manually maintained, browsable and searchable web-based interfaces. Yahoo!

is the most famous subject directory. Yahoo! has several subject headings. A subject directory contains information that is organised into categories and subcategories or topics and subtopics. Like a search engine, one can search a subject directory for all entries that contain a particular set of keywords. Although directories can be searched using keywords, it is often as easy to click on a category, and then click through specific subdirectories until one finds the desired subject. Most web directories are designed to be searched by browsing the subject categories. Directories differ from indexes; however, in the way they organise information. Web indexes simply compile a growing number of records while directories, in contrast, organise information into groups of related records. Besides, Yahoo!, the best-known directories include Open Directory Project (Dmoz.org) and LookSmart.

Directories contain fewer resources than search engine databases. It is because resources in a directory are manually selected, maintained and updated. This, in turn, can be to the advantage of users especially if he or she is searching for a general topic. The directories, therefore, increase the likelihood of retrieving relevant results and increase the possibility of finding high quality, reliable websites.

The directories also have some drawbacks. It is possible that items with similar topics may be placed under two different subject categories in a directory. Directories may not be as current as search engine databases because while the search engines update their databases automatically using robots or spiders, directories are updated manually after new entries are selected, rated and categorised. The directories may also miss out on some of important resources, since resources are selected manually.



Fig. 13.6: Yahoo, the Most Popular Web Directory

13.5.4 Hybrid Search Engines

In the early days of the web, a search engine presented either crawler-based results or human-powered listings. Today, it is very common for search engines to present both types of results. Usually, a hybrid search engine favours one type of listings over the other. For example, MSN Search is more likely to present human-powered listings from LookSmart. However, it does also present crawler-based results (as provided by Inktomi), especially for more obscure queries. There are also many directories that attempt to organise the Internet by subject, and, there are many search engines that combine directory and keyword search capabilities.

13.5.5 Subject Gateways or Subject Portals

Subject gateways, variably called as meta resources, subject-based information gateways (SBIGs), subject-based gateways, subject index gateways, virtual libraries, clearing houses, subject trees, pathfinders and guide to Internet resources are facilities that allow easier

access to network-based resources in a defined subject area. For consistency of terminology, this Unit would use 'subject portals' in preference to other prevailing terminology. Subject portals redirect a user to the holders of the original digital material. It may provide its own indexing and search services or it may combine original resources from a number of different providers. The subject portals restrict their operation to providing linkages to electronic resources hosted on other servers.

A subject portal can be defined as an organised and structured guide to Internet-based electronic information resources that are carefully selected after a predefined process of evaluation and filtration in a subject area or specialty. Subject portals are often independent web-sites or part of an institution or library's website that serve as a guide to Internet resources considered appropriate for their target audiences. A subject portal site that is a part of an institutional website or the library's website, may include resources that are on subscription by the parent organisation and are accessible for free to all. A subject portal may also be built by a commercial enterprise that is accessible free of charge upto the bibliographic level. However, a user may be required to pay if he / she wishes to access the full-text. Home pages of all the major education and research institutions, especially in the developed world, provide an organised and structured guide to electronic resources available on the Internet. Some of the examples of subject portals are LibrarySpot, Librarian's Index to Internet, Argus Clearing House, BIOME and BUBL.

Self Check Exercise

- 6) Name different types of robots used by the primary search engines?
- 7) What are subject gateways? How are they different from Meta search engines?

Note : i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

13.6 CHOOSING A SEARCH ENGINE

The criteria given below may be used to assess the usefulness of a search engine:



Fig. 13.7: LookSmart Search Engines with Instructions For Using Boolean Operators & Phrase Searching

13.6.1 Ease of Use

Most search engines provide a single search window for input of search terms. Search engines query its database for search terms keyed-in by the user. Some search engines have the capabilities of understanding natural language. In a search engine capable of understanding natural language, a user may pose the question exacting as he / she would ask it at the reference desk in a library, i.e., to find out the name of the President of India, he / she may type "Who is the President of India" in the query box. 'Advanced search' in most search engines provides multiple drop-down menu for selection of operators for combining search terms and specifying other parameters that provide a degree of control over the search. Quite often the same controls are available from the standard search window using rather less user-friendly symbols such as AND, OR, NOT, NEAR, (), +, -, ",", fieldname:, etc.

13.6.2 Comprehensiveness

Several well-established search services index hundreds of millions of web documents. Generally, they index full texts of documents. None of the existing search engine indexes complete spectrum of information available on 233 million servers hosted on the Internet. As the contents available on the web increases, the proportion indexed would decrease. Currently, the maximum coverage by any one search engine is estimated to be 34%.

13.6.3 Quality of Content

Search engines are increasingly paying attention to techniques that address the issue quality of content. Relevant and qualitative search results may be given higher weightage than speed and size of search engines. Additional processing, be it human or machine, is used to sift, sort and add value to search results. Examples of such processing are:

- Subjective evaluation and rating by a human classifier;
- Automatic checking of how often a resource is linked to others; and
- Automatic processing of data on previous accesses to the resource.

The results of such processing, combined with matching of search terms, generate a relevance score to be used in the ranking of search results. Currency of the information in the database, particularly working links, is another aspect of quality. A number of search engines revisit URLs at a variable rate depending on how often changes are made to the page.

13.6.4 Control Over the Search

In case of complex queries, the capability for specifying search parameters in detail becomes important. A user should be able to combine multiple search terms with ease. Most search engines usually allow users to specify how search terms are combined, either by typing the search string using the Boolean terms AND, OR and NOT into the search window, or giving equivalent functionality via drop-down menus. The use of parentheses in nesting Boolean search combinations provides further level of control. Search techniques described in the next section of this Unit elaborates on various techniques and how they control the search results.

13.6.5 Flexibility in Searching

Advanced search features in some of the search engines provide following flexibilities:

- Provision to restrict the search to specific fields, (i.e., title, description, keywords, links, body, etc.)
- Provision to restrict a search to a specified time period;
- Provision to search for similar documents (if a user finds something useful); and
- Provision to search within the results of a previous search.

13.6.6 Assessment of Relevance

A good search engine should take quality of resources into account while ranking search results. Search engines display the search results either using the frequency of keywords used in the web page, positioning of keywords on the web page (title, description, body, etc.) and frequency of inbound-links from other websites. Irrespective of ranking mechanism, a user would prefer to see document relevant to his / her search in the first few search results.

13.6.7 Informative Presentation of Results

Search engines should record the number of hits found for a search. A user would prefer to get enough information to judge the usefulness of a link before following it. Most search engines take some portion of the text to provide an abstract. Dates are often useful. A user would also like an option to specify alternative criteria for ranking, for example by location, date, etc. A user would prefer live links, i.e., the search engine should validate links and update its database frequently.

13.7 SEARCHING THE WEB: SEARCH TECHNIQUES

13.7.1 Searching Subject Directories

Browsing a directory is a simple matter of following the links for a given topic of interest. Searching either a directory or the portion of the web that a search engine covers works in a similar way in almost all the search engines and directories.

As discussed before, a web directory (like Open Directory Project, Google Directory or Yahoo) is a subject catalogue that organises the web into major topics and sub-topics such as Arts, Business and Economy, Computers and Internet, Education, Entertainment, Government, Health, News, Recreation, Reference, Regional, Science, Social Science, Society and Culture. Under each topic, subtopics are listed, and under each subtopic further subtopics or links to resources are listed.

To find about 'Web Page Designing' from Yahoo, for example, select Computers and Internet that lists Wide World Web (WWW) as an additional category. Click on it to find another list of additional category that includes 'Site Development', clicking at site development leads to several links that are pertinent to the search topic.

The web directory is a better source of information than a search engine if a user is clear about his or her query. Directories generally do not give as many references as a search engine would, but the references retrieved through a directory are likely to be relevant.

Web directories are usually equipped with their own search engines that allow a user to search through their indices for the information they need. In fact, search engines and web directories are being integrated in an interesting way. For example, a search conducted on Google search engine and one of the results happens to be in the Google's Directory (which is based on the Dmoz directory), Google will offer a link to that section of the directory. Similarly, results for a search conducted in the Google directory, are arranged according to PageRank, which is Google's all-important measure of 'link popularity'.

13.7.2 Searching Search Engines

As mentioned before, search engines deploy automated software programs known as spiders or robots that survey the web and build their databases of keywords, titles, URLs or full-text. web documents retrieved by these spiders or robots are culled together and analysed by these programs. Data collected from each web page are then added to the search engine index. When a user makes a query to the search engine, the query is checked against the search engine's index of all the web pages and relevant documents

with their URLs are returned as hits. These hits are ranked in order of relevance with the best results on the top.

Most search engines offer two types of interfaces to search their databases, i.e., basic search and advanced search. In a basic search, a user just keys-in his / her search term without sifting through pull-down menus for additional options. Full-featured search engines 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. Option for 'Enhance Search' must be used to obtain more relevant and specific search results.

Basic Search

Most search engines offer a dialog box, pane or a dialog line where search terms can be keyed-in followed by options to either submit or clear the search. Most search engines query their database using keywords. A user enters a word or words called "keywords or search terms" that he or she would like to search. The search engine then looks through its indexes in the database for matches. It might look in the title, description or entire text of a web page. After conducting the search, a list of sites that matches the requested search terms are produced as results. Some search engines also give an option to search within the results retrieved in preceding search.

Advanced Search or Refining Your Search

Different search engines have different methods of refining queries. The best way to learn them is to read the help files on the search engine sites and practice. Options for advanced search differ from one search engine to another, but some of the common features include ability to search on more than one word, to confine the search to a specified field and to exclude words that are not required in a search by the user. A user may also search for proper names, phrases, and on words that are found within a certain proximity to other search terms. Some search engines also offer some degree of customisation of results. Several search engines allow the use of Boolean operators or signs like '+' or '-' to refine the search.

Boolean Operators

Many search engines allow use of 'AND', 'OR' and 'NOT' to narrow or broaden a request. Boolean operators allow us to connect the concepts of our search query in a meaningful way so that a user can retrieve relevant search results.

Operator 'AND'

Operator 'AND' is used to narrow the search results to two terms combined with operator 'AND'. The example given below illustrates the search statement: Library **and** Automation. The results will consist of websites (or other Internet information) that contain both words, i.e., Library and Automation. A user may also use '+' sign in place of AND in most of the search engines.

AND

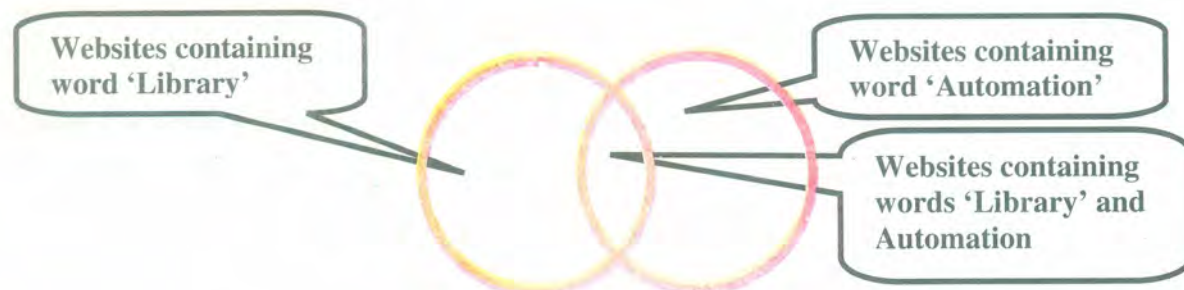


Fig. 13.8: Using Logical Operator AND

Operator 'OR'

The operator 'OR' is used to broaden the search results to any of the two terms combined with operator 'OR'. It retrieves pages that include any of the terms given in search query. The example given below illustrates the search statement: Library **OR** Automation. The operator 'OR' is very helpful for combining synonyms. It is similar to the common search engine option, "SEARCH ANY WORD".

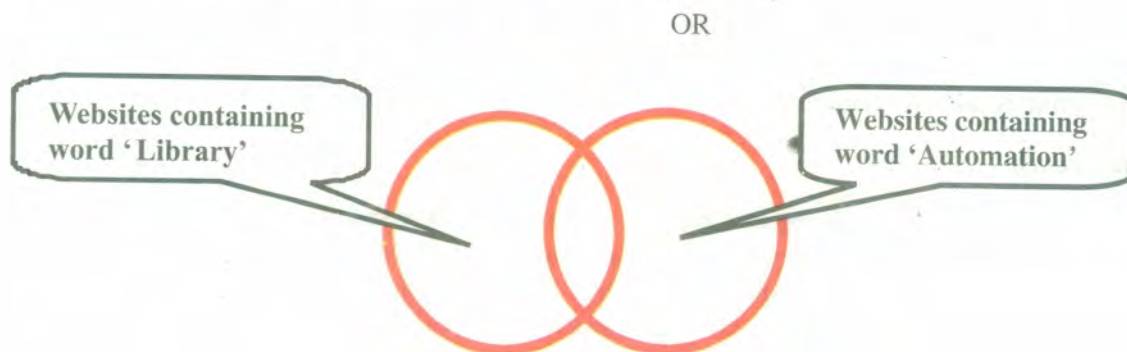


Fig.13.9: Using Logical Operator OR

Operator 'NOT'

The operator NOT is used to eliminate terms from the search results. It is used to *exclude* a particular term from the search results given after 'NOT'. It retrieves pages that exclude a word specified after 'NOT'. The example given below illustrates the search statement: Library **AND** Automation **NOT** India. A user may also use '-' sign in place of NOT in most of the search engines.

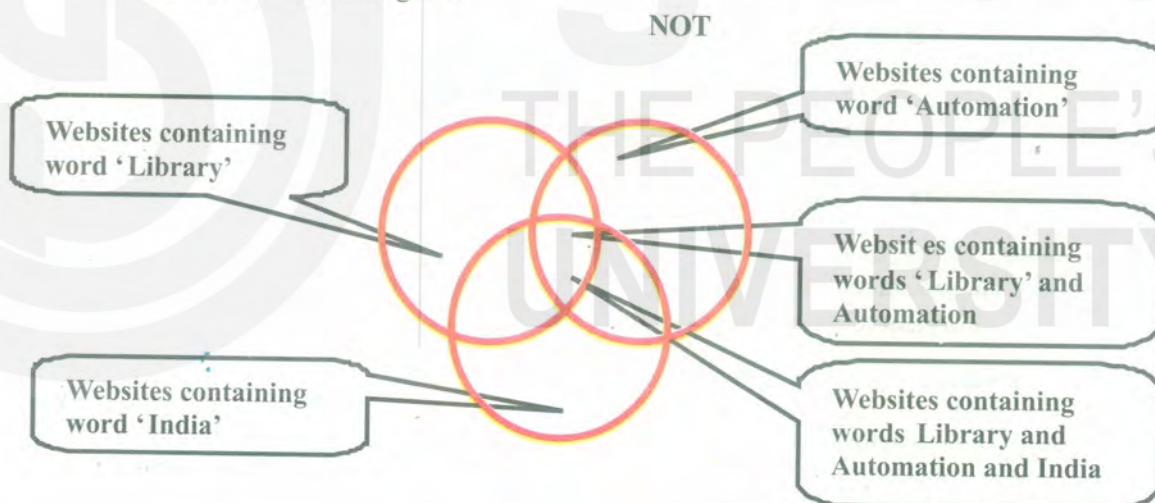


Fig. 13.10: Using Logical Operator NOT

Operator 'NOT' should be used with caution, as it might result in elimination of relevant results from a search. This will happen if there are documents that contain both the concepts being searched on and the concept that is being eliminated.

Phrase Searching

The ability to query on phrases is very important in a search engine. A phrase is a group of words that must appear next to each other in a specified order. Phrase searches are especially useful when searching the famous sayings or proper names. Most search engines support this feature. It can be used when the search terms appear in an exact order. To indicate a phrase, surround it in double quotation marks. For example the following searches can be made as phrases:

Phrase searching is one of best search features that can be used to increase the chance of retrieving relevant results.

Proximity Searching

Proximity operators are used to specify the relative location of words in a document. These operators facilitate searching for words that must be in the same phrase, paragraph, or sentence in a record. Proximity operators help us to search for words within a certain distance of one another in databases. For example, a search may require that two concepts be in the same sentence but not necessarily next to each other, as in a phrase. One such operator is NEAR which means that the terms that are entered should be within a certain number of words to each other. Typically, the distance between two concepts can be 10-25 words. NEAR allows the terms to be in any order. Different search engines may use different proximity operators. The symbols generally used in this type of search are ‘w’ (with/within) and ‘n’ (near). For example a search for “library ‘near’ automation” would retrieve documents containing ‘Library automation’ and ‘Automation of Library’.

Another operator used is ‘followed by’ which means that one term must follow the other. ADJ (adjacent) serves the same function. A search engine that allows search on phrases essentially uses the same method, i.e., determining adjacency of keywords.

Parentheses

Most search engines permit the use of parentheses to group related terms. This is particularly useful for clustering synonyms or for searching specific terms together before other terms are searched. Parentheses may be used in combination with other search techniques. The example below indicates how to use this feature:

(Library Computerisation or Library ‘near’ Automation) and India

Truncation and Wildcards

Truncation is a technique that facilitates search for multiple endings of a word. It is also called stemming. Most search engines that support this feature use certain symbols (called ‘wild card’) such as *, ? or # at the end of the word root to indicate a truncated search. The following examples show how to construct a search using truncation:

class* (retrieves class, classification, classify, class no.)

catalog* (retrieves catalog, catalogs, cataloguing, catalogue, catalogues, catalogued)

It is generally a good idea to truncate longer word roots. Truncating shorter roots, such as cat*, will usually result in several hits, with many irrelevant results.

Wild cards can also be used for internal truncation. Internal truncation is generally used to search for words that may differ by one or two letters within a word. Some instances are words with British and American spellings. The examples given below indicate how to use the wildcard feature:

wom*n (retrieves woman or women)

colo*r (retrieves colour or color)

hono*r (retrieves honour or honor)

Please note that wild cards (* in the example given above) may vary from search engines to search engines.

Case Sensitivity

Case sensitivity refers to the ability of search engines to distinguish between upper and lowercase letters. Most search engines are not case sensitive and will simply read all letters as lowercase. Others may distinguish between the word 'aids' and the disease 'AIDS'. Using lower case is advised, because lower case will always retrieve upper case as well.

Field Searching

Web pages are made up of different parts or 'fields'. Several search engines can limit a search to a specific area of a web page. This technique helps to increase the relevance of search results. The help section of search engines that support field searching would describe which fields can be searched. The actual field names may differ among search engines. The following fields are commonly supported:

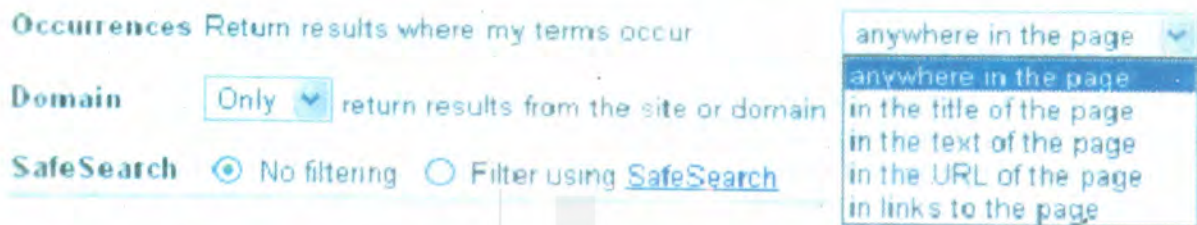


Fig. 13.11: Pull-down Menu in Advanced Google Search for Limiting Search to Specific Field in Document

Title Search: Finds pages that contain the given term in the page title (which appears in the title bar of most browsers).

Command	Supported By	Examples
title:	AltaVista, AllTheWeb	title: automation
in title:/	Google	in title: automation
allintitle:		allintitle: automation

URL Search: Looks for words that are parts of the URL (web address). For example: url: niscair, would look for URL having niscair in it.

Command	Supported By	Examples
url.all:	AllTheWeb	url.all:niscair
allinurl:/inurl:	Google	allinurl:niscair
		inurl:niscair

Link Search: Used for finding all the web pages that point to the given url.

Command	Supported By	Examples
link:	AltaVista, Google, Northern Light	link:www.iitd.ac.in
linkdomain:	Inktomi, MSN	linkdomain:www.iitd.ac.in

Site Search: A user can restrict his or her query search within a domain. This can be helpful in obtaining highly relevant results from a specific site.

Command	Supported By	Examples
host:	AltaVista	indest host:www.iitd.ac.in
site:	Excite, Google	indest site:www.iitd.ac.in
domain:	Inktomi (HotBot)	indest site:www.iitd.ac.in

File Types Search: Some search engines allow the user to restrict his or her search to a specified file format.

Command	Supported By	Examples
Filetype	Google	file type:pdf



Fig. 13.12: Pull-down Menu in Advanced Search of Google to Restrict a Search to a File Type

Definition Search: This command allows searching for definition for a word or phrase.

Command	Supported By	Examples
define:	Google	define:search engine

In addition to the above, different search engines provide many other features.

Concept Searching

Unlike keyword search systems, concept-based search systems try to determine what a user means. In the best circumstances, a concept-based search returns hits on documents that are 'about' the subject / theme that is being explored, even if the words in the document do not precisely match the query.

There are various methods of building up an algorithm for concept searching, some of them are highly complex, relying on sophisticated linguistic and artificial intelligence theories.

Excite, for example, uses a numerical approach. Excite's software interprets meaning by calculating the frequency with which certain important words appear. When several words or phrases that are tagged to signal a particular concept appear close to each other in a text, the search engine concludes, by statistical analysis that the piece is 'about' a certain subject.

For example, the word heart, when used in the medical / health context, is likely to appear with such words as coronary, artery, lung, stroke, cholesterol, pump, blood, attack and arteriosclerosis. If the word heart appears in a document with others words such as flowers, candy, love, passion and valentine, a very different context is established, and a concept-oriented search engine returns hits on the subject of romance. Concept-based searching is not supported by all search engines.

Natural Language Searching

The natural language searching feature allows a user to search in the same language as it is spoken. Suppose a user wants to know who was Jawaharlal Nehru. His / her query would be: "Who was Jawaharlal Nehru?". The search engine that supports this technique retrieves relevant web pages that would answer this question.

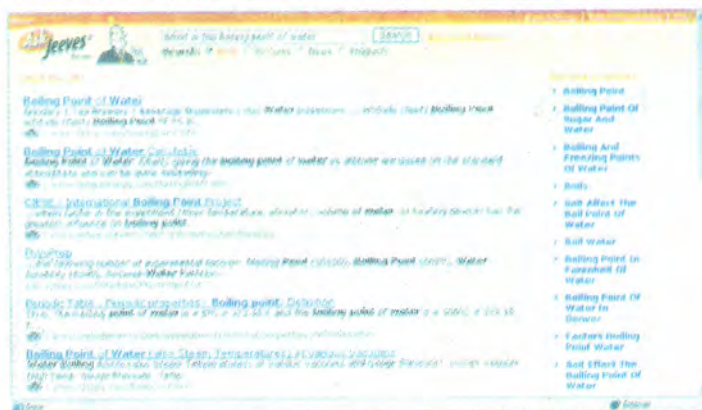


Fig. 13.13: Results of a Search "What is the Boiling Point of Water" on 'Ask Jeeves'

13.8 SEARCH RESULTS

After entering the search terms, once a user clicks at 'submit' or 'search' button, search engines usually provide a list of best matches for the search term keyed-in by the user. The search softwares are designed to rank the sites in order of relevance. Different search engines use different algorithms to measure relevance. Most search engines use frequency of keywords and their positioning in a document to determine its relevancy, considering that if the keywords appear more frequently and appear early in the document, or in the headers, it is likely to be more relevant. Frequency with which a document is linked to other documents on the web is considered as a measure of its popularity and relevance by the Google search engine, and search results are ranked and arranged accordingly.

13.8.1 Result Lists

Generally, search engines display 10 to 15 search results in a page. At the bottom of the list there are options to click for next page and / or options to customize number of hits that should appear on a page. The software may have a limit to the number of sites it can retrieve. Most search engines allow users to fine-tune the search, either through advanced search or by searching within the search results. Observations reveal that if relevant results are not found in the first 50 hits, chances are that there may be none and a user may have to start the search process all over again.

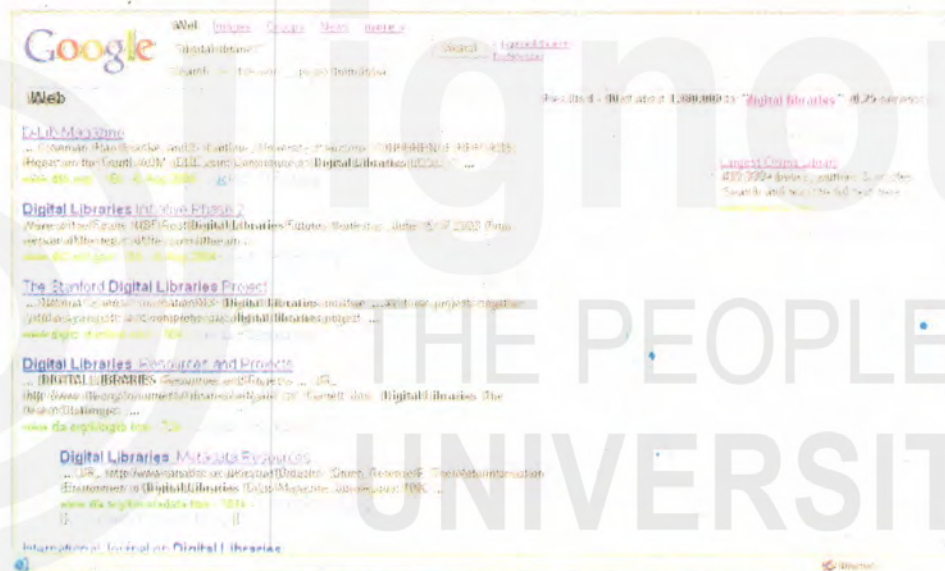


Fig. 13.14: Search Results for a Search on Digital Libraries in Google

13.8.2 Result Descriptions

Most search engines offer search results with descriptions of sites to help the users to determine whether the link is worth following. Some search engines include the first few paragraphs of a document along with the URL of the site. Many search engines simply give listing of URLs with very little description.

13.8.3 Relevance Ranking

Once a user finishes his search, the search engine retrieves results and displays the hits. How a search engine organises its findings varies greatly from search engine to search engine. Most of the search engines return results with confidence or relevancy rankings. In other words, they list the hits according to how closely they think the results match the query. In spite of ranking methods used and advanced search facilities offered, users find thousands of websites as a result of their searches, many of them seem completely irrelevant.

Most search engines use *frequency* of keywords on a web page as a primary way of determining whether a document is relevant or not. A search engine will analyse how

often keywords appear in relation to other words in a web page. Those with a higher frequency are often deemed more relevant than the others. Some search engines consider both the frequency and the *positioning* of keywords to determine its relevancy, considering that if keywords appear early in the document, or in the headers, it is likely that the document is more relevant. For example, one method is to rank hits according to how many times keywords appear and in which fields they appear (i.e., in headers, titles or in plain text). *Links* given by other websites is a popular method used for ranking of websites considering the fact that there are several other users that find the site useful and important.

As far as the user is concerned, relevance ranking is critical, and becomes more so as the sheer volume of information on the web grows. The user neither has patience nor time to go through scores of hits to determine which hyperlinks he / she should actually explore. A user would naturally prefer a search engine that provides more relevant sites in its first few results.

13.9 META TAGS

Meta tags provide direction and information for search engine robots. They provide a variety of identifying information for the search engines to index the site. Words, terms, names, etc., encircled in the meta tags provide cataloguing information to the search engines enabling them to index the site. There are several meta tags. The ones that concern search engines are author, contributor, title, description and keywords, etc. Meta tags are not viewable to Internet users unless they examine the source code of the page. Search engines often pick up the description when showing search results. Meta tags that can affect search engine rankings are keywords and description. However, most search engines do not consider meta keywords for positioning.

```
<html>
<head>
<title>UCLA Homepage</title>
<!-- This page was auto-created on Sunday, August 8, 2004 at
9:30:4 -->
<meta http-equiv="Content-Type" content="text/html; charset=iso-
8859-1">
<meta http-equiv="Author" content="pirco">
<meta name="description" content="UCLA (University of Califor-
nia, Los Angeles) home page. UCLA is the largest UC campus in
terms of enrollment, and one of the few public research universities
located in a major city.">
<meta name="keywords" content="UCLA, University of Califor-
nia Los Angeles, UCLA Bruins">
```

Fig. 13.15: Meta Data as Given in the Website of the University of California at Los Angeles (www.ucla.edu)

Different search engines look at meta tags in different ways. Some rely heavily on meta tags, others do not use them at all. Google, currently the most popular search engine, does not index the keywords in meta tags. The 'title' and the 'description' meta tags are considered important for websites, since several major search engines use them in their indices. Use of relevant keywords in title on different pages is important for the website.

Unless the author of the web document specifies the keywords for his / her document, it is up to the search engine to determine them. Essentially, this means that search engines pull out and index words that appear to be significant. Since search engines are software programs, not rational human beings, they work according to the rules established by their designers for what words are usually important in a broad range of documents. The title of a page, for example, usually gives useful information about the subject of the page. Words that are mentioned at the beginning of a document are given more weightage by most search engines. The same goes for words that are repeated several times throughout the document.

13.10 SEARCH ENGINES: EVALUATION

Evaluation of a search engine is the process of identifying and collecting the data about the search engine and establishing criteria by which its success can be assessed. Evaluation of a search engine should determine both its quality and the degree to which it has accomplished its stated goals and objectives.

As discussed previously, search engines consist of three parts: i) robot or web crawler; ii) a database of web documents, iii) an agent or a search engine operating on that database as well as a series of programs that enable display of search results. Joe Barker, University of California has identified the following criteria for evaluation of search engines for the three different components:

13.10.1 Database of Web Documents

Size of Database and Method of Compilation

- i) How many documents does the search engine claim to have?
- ii) How much of the total web can be searched using the search engine?
- iii) How is the index compiled (collection method)? Automated or human input?

Currency or Up-to-date

- i) Search engine databases consist of copies of web pages and other documents that were made when their crawlers or spiders last visited each site. How often is the database refreshed to find new pages?
- ii) Does the search engine claim to revisit sites that add to or change their information? If so, how often and how many sites?
- iii) How soon are pages included in the index after they are collected either by spiders or through submissions?
- iv) What is the number of pages crawled per day? How long does it take for the search engine to update its entire index?
- v) Is there evidence of link checking (i.e., dead or out-of-date links)?

Indexing

- i) Are there any provisions for use of controlled vocabulary?
- ii) Does it use a stop word list? How extensive it is? Is it documented as to what it identifies as a stop word?

Coverage

- i) Types of resources indexed in the database of the search engine (ftp, www, newsgroups, etc.)
- ii) Are there any special criteria for inclusion?

Completeness of Text

- i) Is the database really 'full text' or only parts of the pages are collected?
- ii) What elements (e.g., title, keywords, descriptors, body) of a page are indexed?
- iii) Is every word indexed?

Types of Documents Covered

- i) All search engines search web pages. Do they also have extensive PDF, Word, Excel, PowerPoint, and other formats like WordPerfect?
- ii) Are they full-text searchable?

- i) How fast is it?
- ii) How consistent is it? Do you get different results at different times?

13.10.2 Capabilities of a Search Engine

- i) Does the search engine spider have access to password-protected sites?
- ii) Is the spider able to follow frame links and image maps?
- iii) Where cannot the search engine reach?
- iv) All search engines let you enter some keywords and search on them. What happens inside?
- v) Can you limit the ways that will increase your chances of finding what you are looking for?

Basic Search Options and Limitations

- i) Is automatic default of AND assumed between words?
- ii) Does search engine accept “ ” to search phrases?
- iii) Is there an easy way to allow for synonyms and equivalent terms (OR searching)?

Advanced Search Options and Limitations:

- i) Can you restrict your search terms to specific fields, such as the document title?
- ii) Can you restrict some words to certain fields and some others to other fields?
- iii) Can you restrict to documents only from a certain domain (org, edu, gov, etc.)?
- iv) Can it be limited to more than one or only one?
- v) Can you limit by type of document (PDF or excel, etc.)? More than one type?
- vi) Can you limit by language?
- vii) How reliably and easily can you limit to the date last updated?

General Limitations and Features

- i) What do you have to do to make it search on common or stop words?
- ii) What is the maximum limit on search terms or on search complexity?
- iii) Does it have the ability to search within previous results?
- iv) Can you count on consistent results from search to search and from day to day?
- v) Can you customise the search or display?
- vi) Is there a ‘family’ filter? Does it work well? Is it easy to turn it on or off?

13.10.3 Results Display

- i) All search engines return a list of results it ‘thinks’ are relevant. How relevant the results actually are?

Ranking

- i) Do pages with search terms juxtaposed (like a phrase) rank highest?
- ii) Do you get pages with only some of your words, perhaps in addition to pages with them all?

Display

- i) Are your keywords highlighted in context, showing excerpts from the web pages, which caused the match?
- ii) Are some other excerpts from the pages displayed?

Collapse Pages from the Same Site

- i) Does the display show only one or a few pages from the same site? Does it show the one(s) with terms that are searched?
- ii) How easy is it to see all the results from the same site?
- iii) Can this be changed and saved as your preferred search method?

13.11 IMPORTANT SEARCH ENGINES

There are hundreds of search engines on the Internet, and more are being created every day. As shown in Figure 13.16, shows first the 15 search destinations. Details about a few important search engines are given below:

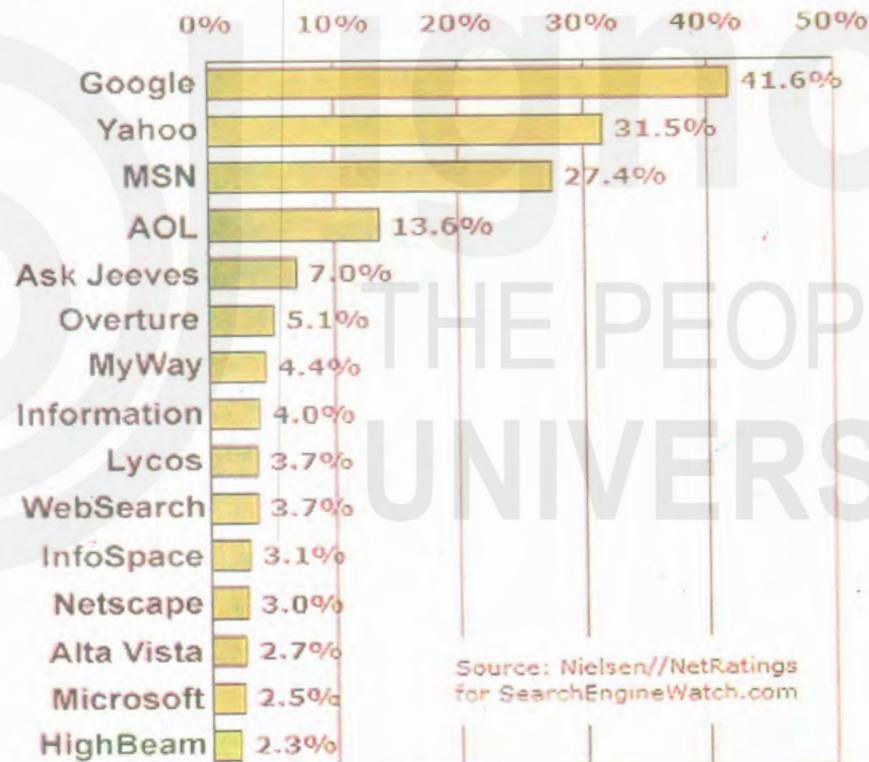


Fig. 13.16: Top 15 Search Destinations (SearchEngineWatch.com)

13.11.1 Primary Search Engines

Google (<http://www.google.com/>)

Google is not only new but next generation search engine that consists of 4.3 billion pages. Developed by several Stanford PhD students, it has followed Yahoo, Excite, and WhoWhere in moving from a student project to a commercial site. Its relevance ranking uses two factors not generally included in search engine rankings, i.e., the number of links to the page from elsewhere and the 'importance' of the pages that link to it. Other ranking factors are the number of hits on the search words in the title and the text and the proximity of search terms to each other.



Fig. 13.17: Google Search Interface

Google has become the pre-eminent web search engine for many, since it was launched in 1999. It has made its mark with its relevance ranking based on link analysis, cached pages, and aggressive growth. Since its beta release, it has had phrase searching. In June 2000 it announced a database of over 560 million pages, which grew to over 600 million by the end of 2000 and then to 1.5 billion by December 2001. The 2+ billion visits reported on their home page as of April 2002, includes indexed pages, unindexed URLs, and other file formats. By November 2002, they moved their claim up to 3 billion, and in July 2004 it went to 4.26 billion. The biggest strength of Google is the size of its database and link-based relevance ranking.

Its database, besides web pages, includes additional file types such as PDF, .ps, .doc, .xls, .txt, .ppt, .rtf, .asp, .wpd and more. The Google also has a database of images, Usenet News group, paid advertisements that are usually shown on the right side under 'Sponsored Links' and web-based news sites.

Google supports Boolean searching, proximity searching, field searching, limit to file type, language and domain, character searching, numbers and number range searching. Results are sorted by relevance that is determined by links from other pages with a weightage given to authoritative sites. Pages are also clustered by site. Only two pages per site are displayed, with the second indented. The display includes the title, URL, a brief extract showing the text near the search terms, the file size, and for many hits a link to a cached copy of the page. This cached copy is from Google's index and may be older than the version currently available on the web. The cached copy will display highlighted search terms. If more than one search term is used, each has a different colour highlighting. The default output is 10 hits per screen, but the searcher can also choose 20, 30, 50, or 100 hits at a time on the preferences page.

MSN Search (<http://search.msn.com/>)

MSN Search is one of the search engines for the MSN portal site. It uses an Inktomi database. The basic search screen only shows a few options, but by choosing the Advanced Search link, the full range of search features is displayed. This review discusses the full set of options, some of which are only available in the advanced search. MSN Search uses LookSmart for its directory and Inktomi for its search engine database. Its sponsored sites (ads) are from Overture. MSN Featured Sites and Directory results come first from the basic search screen. The advanced search only displays Inktomi results.

MSN search supports Boolean searching, proximity searching, truncation, field searching, restriction to media type and file types. Results are sorted by relevance. Unique among

the Inktomi group, the MSN advanced search also provides an option to sort “equally relevant results” by date, depth, or title. Note that this will only sort those records which have exactly the same relevance ranking score. The display includes the title, a brief summary, and the URL. MSN Search displays 15 records at a time, by default. However, users can request displays of 10, 15, 20, or 50 records at a time in the advanced search or change through the preferences. The advanced search also has an option to exclude the summary.

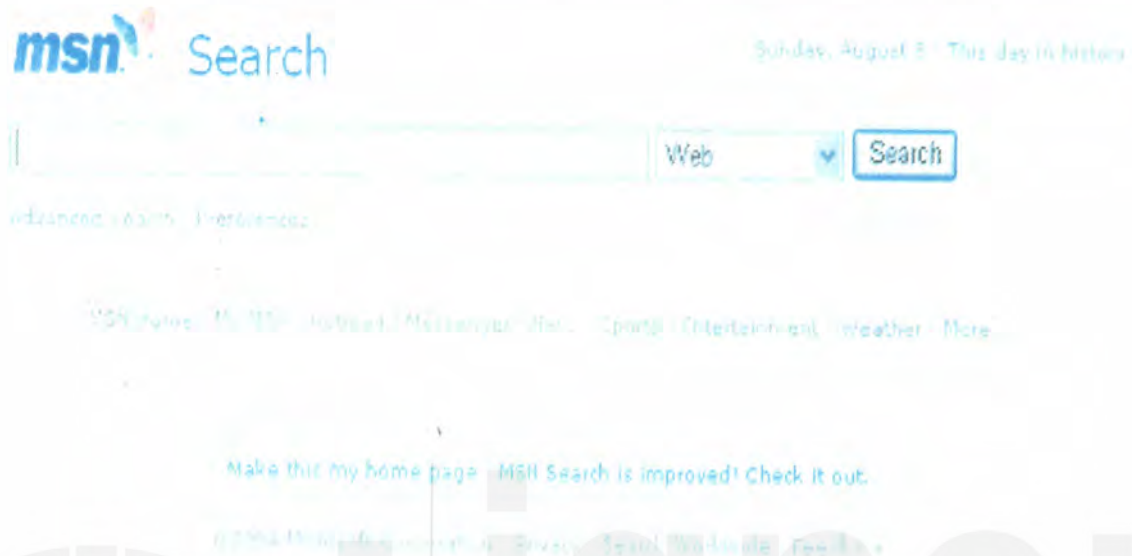


Fig. 13.18: MSN Search Engine

Teoma

(<http://teoma.com/>)

Launched in Spring 2001, this new search engine has built its own database and offers some unique search features. It was bought by Ask Jeeves in Sept. 2001. It lacks full Boolean and other advanced search features, but it has more recently expanded and improved its search features and added an advanced search. While Teoma results can show up in three separate sections, it has only one database of indexed web pages. It may also include paid ad results (from Google's AdWords database) under the heading of 'Sponsored Links'. No additional databases or portal features are directly available. Ask Jeeves switched to Teoma instead of Direct Hit in Jan. 2002 for the search engine results. Teoma supports Boolean searching, proximity searching, field searching and searches restricted to field in a limited way.

WiseNut

(<http://wisenut.com/>)

Launched in beta in 2001, this new search engine has built its own large database. Like Teoma, it features some automatically generated, semantically related searches, known as WiseGuide categories. It lacks full Boolean and most advanced search features. WiseNut has one single database of indexed web pages. No additional databases or portal features are available as of now. WiseNut does have one partner site, Korea WiseNut. It supports proximity searching and search limits to languages. By default, sites are sorted in order of perceived relevance. There is no option for sorting alphabetically, by site, or by date. Results display title, 1 or 2 lines as it is, number of links to other matches from the same site and the URL. WiseNut clusters results by site, but aids the searcher by giving the number of other matches from the same site. Above the search results, WiseNut displays the query, total number of results, the WiseGuide categories, and then the first 10 results.

WiseNut offers a personalisation capability that allows users to control display of the number of records with or without summary, turn site clustering on or off, choose encoding scheme and turn WiseGuide categories on or off.

Fast Search and Transfer (<http://www.fastsearch.com/>)

Fast Search & Transfer (FAST) created the search engine and database for AlltheWeb and Lycos. The web search unit of FAST was bought by Overture in February 2003. Launched in 1999, Fast Search and Transfer, offered their Fast search engine at www.alltheweb.com. The first major expansion came in January 2000 with the appearance of their database on the Lycos site. Starting at 80 million URLs, Fast grew to 200 million by August 1999 and 300 million in January 2000. They passed 2 billion in June 2002. The search engine plans to index all the public, indexable web.

The FAST Search interface consists merely of a navigation bar and a search entry box and a number of links at the page bottom. There is very little in the way of advertising on any of the site's pages. The Advanced Search interface is different only in so far as it provides additional options for the searcher.

Alta Vista (<http://www.altavista.com/>)

AltaVista, a business of Overture Services, Inc., is a leading provider of search services and technology. AltaVista, which means "a view from above", was launched in 1995 when scientists at Digital Equipment Corporation's Research lab in Palo Alto, CA, devised a way to store every word of every HTML page on the Internet in a fast, searchable index. This led to AltaVista's development as the first searchable, full-text database on the World Wide Web. Other notable AltaVista inventions include the first-ever multi-lingual search capability on the Internet and the first search technology to support Chinese, Japanese and Korean languages.

The Alta Vista was the first Internet search engine to launch image, audio and video search capabilities. It was the first search engine that offered advanced search features and capabilities like multimedia search, translation, language recognition and specialty search. The Alta Vista allows users to perform simple or complex searches and has speedy retrieval times and well-developed robot technology. If no connector is used in the search, the default is 'or.' Truncation is possible. A user can limit his / her search to a field, URL, title and links. The link search retrieves pages where at least one link represented on that page matches the search query. Advanced searching is also available by using Boolean operators and adjacency symbols. The search engine supports proximity operators and parentheses for nesting.

Web pages are evaluated for relevance, i.e., its ranking system is not as effective as that of other search engines because it indexes any and all references to a search term, no matter how far off it may be from the query's intent. Its search engine does not allow truncation as other search engines do, which means that searches are performed using the exact phrase, i.e., plurals and other forms of words are left out. However, if a document is found in the search, a user can be sure that his / her search terms are somewhere in it. Alta Vista also provides dates in its results list. Although a user can refine his / her search by using the Power Search option, Alta Vista does not have as much on-screen help as other search engines. Alta Vista allows a user to bookmark results, making future site visits much easier.

Go.com (<http://www.go.com/>)

Formerly InfoSeek, Go.com is owned by the Walt Disney Internet Group (WDIG) and powered by the Google search engine. It is a free directory and keyword searchable service. A user can browse the directory for various topics and subtopics or look for specific information through its search interface. Go.com indexes over 1 million web pages. It also indexes Usenet newsgroups, FTP and Gopher sites, e-mail addresses, and Frequently Asked Questions lists. Search features are similar to Google Advanced Search and include Boolean search with AND, OR and NOT. It supports phrase searching and proximity searching. Results are ranked by relevancy and include that ranking, a link to the site of the information, the URL of the site, the size of the document, some description of the document, and a link to similar pages. A user can bookmark results, making return visits to the sites much easier.

Lycos (<http://www.lycos.com/>)

Back in December of 1995, Lycos claimed to have indexed 92% of the web. Now, it claims to be the only complete guide to the Internet. Lycos graduated from being simply a keyword searchable index to adding a directory, which goes by the name of A2Z. Lycos also provides a service called Point, which provides reviews and ratings of the top 5% of all the Internet sites they index. Lycos searches every word in a website. The advanced search features include Boolean searches through a pull-down menu, i.e., any of the word, all of the word and none of the word. Searches can be restricted to desired locations in web pages, i.e., title, description, body or the URL. A search can also be restricted to a language. A user can also choose the level of relevancy of the search. Display options range from showing 10-40 results per page in either standard, summary or detailed form. Standard display includes a link to the document, the relevancy ranking, an outline, an abstract, the URL, and the size of the document.

Excite (<http://www.excite.com/>)

This search engine offers two ways of searching: concept or keyword. Many times there are no significant differences between the results of two types of searches. There is no Boolean searching, so trying to find specific information on a topic can be frustrating. The Excite offers a user to choose for a directory search (like Yahoo!) or a keyword search. The entire database is checked and updated weekly by spiders that are sent out on specific missions: One is sent to the What's New sites to compile a database of new URLs. Another is then sent out to bring back the page contents to the Excite database.

Search results cannot be book-marked. URLs are not displayed in the results. It is easy to use and a good place to start for current topics.

13.11.2 Web Directories

Yahoo! (<http://www.yahoo.com/>)

Yahoo! is one of the best known and most popular Internet portals. Originally just a subject directory, it is now a search engine, directory and portal. A search on Yahoo! provides search results that include a few categories from the directory and Inside Yahoo!, followed by sponsored links (ads), and then the bulk of the results based on an Inktomi database. Yahoo is also backed-up with an Image database, Yellow Pages, Products and a News database. Yahoo supports Boolean searching, proximity searching, field searching, limits to language, domain, date, filetype, country and adult content. However, some advanced features such as truncation is not supported.

Results are sorted by a relevance algorithm. Pages are also clustered by site. Only one page per site is displayed. Others are available via "More pages from this site link" after the cached link at the end of the record. Yahoo! provides results in six categories. The first lists of results under web are from the search engine with the page title, a keyword in context extract (or directory description or meta description), the URL, file size, cache link, and a possibly "a More pages from this site link". The second tab is a link to their image database. The Yahoo! directory results are available under the Directory heading. The Yellow Pages tab goes to a Yellow Pages search form. The News tab goes to the Yahoo! News database while the Products tab, goes to the Yahoo! Shopping search.

LookSmart (<http://www.looksmart.com/>)

LookSmart is one of the larger and better-known directories. It is used by AltaVista, AltaVista Australia, Anzwers, MSN Search, and Go2Net. LookSmart has over 2.3 million unique URLs according to the company with 250,000 categories and 200 editors. In addition to the LookSmart database, Direct Hit results are also available via a "Top 10 Most Visited" link. If no hits are found in LookSmart, results from AltaVista are shown, and these may include RealNames results as well. LookSmart Live is a service more than a

database that offers personalised email responses to questions. Some prominent search services that use the LookSmart directory include AltaVista, Anzwers, MSN Search, Netscape, Excite, and Go2Net. LookSmart supports Boolean searching and automatic truncation. It does not support phrase searching, proximity operators, limits to field search, etc.

The sites appear to be sorted randomly, but they may be ordered by relevance. If a LookSmart search fails to find any hits in LookSmart directory, the search is automatically passed on to AltaVista. While this can be handy, it is rarely as effective as a well-structured search run directly on AltaVista. LookSmart first displays categories that contain matches. Then sites are displayed with their title hyperlinked to the URL, a two-line description, and a linked connection to the category in which the site is listed. After the sites, results from AltaVista are displayed.

D MOZ Open Directory Project (<http://dmoz.org/>)

The Open Directory Project, formerly known as NewHoo, is owned by Netscape but is run by volunteer editors. Entries in the Directory are all selected by the editorial team of over 30,800 editors. Many entries come from submissions. Open Directory has over 2,152,000 unique entries divided into over 314,000 categories and subcategories. The Open Directory as seen at <http://dmoz.org/> is only one database. Partners using the Open Directory include AltaVista, Netscape, Lycos, HotBot, and others. D MOZ Open directory supports Boolean searching and truncation. It does not support proximity searching, phrase searching and field searching.



Fig. 13.19: DMOZ Open Directory Project

Categories in DMOZ Open Directory are presented first if they match or sometimes if they contain records that match. Then sites are listed, presumably in order by some relevance score, but the method of ranking is not described. Open Directory first displays up to five categories. Then sites are displayed with their titles hyperlinked to the respective URLs, a brief description, a linked connection to the category in which the site is listed, and a number in parentheses.

Galaxy (<http://www.galaxy.com/>)

The Galaxy is another hierarchical, topically organised search engine. Each topic has its own page in the Galaxy, and each page is organised into many lists. For example, the

Topic Lists page provides links to other Galaxy pages containing specific information about the search topic. Search results consist of a series of indexes from which to choose. One can search on the full word or on the acronym on the Galaxy. Boolean 'and', 'or' and 'not' can be used to refine the search process. The advanced search in Galaxy provides for restricting the search to title, description, body or other locations in a page. A user can also restrict his or her search on a domain name. Galaxy provides for phrase searching and a user can define search depth to medium, shallow, deepest and surface.



Fig. 13.20: Galaxy Search Engine

The Galaxy has a link called "You can add information to this page!". Clicking on it brings up a form which can be used to add references to an existing page, or send comments to Galaxy staff. Each index provides its own results, which are scored according to the frequency with which specified keywords are found.

13.11.3 Meta Search Engines

WebCrawler (<http://www.webcrawler.com/>)

WebCrawler was launched in 1994 and handled its one-millionth query in November of the same year. WebCrawler joined AOL in 1995, and was acquired by Excite in 1997. In 2001, WebCrawler joined the InfoSpace team. WebCrawler uses innovative metasearch technology to search the Internet's top search engines, including Google, Yahoo, Ask Jeeves, About, Teoma, FindWhat, LookSmart, and many more. With one single click, WebCrawler searches the best results from the combined pool of the world's leading search engines - instead of results from only one single search engine. WebCrawler makes it easy to refine the search so that a user can find the most meaningful results right away.

WebCrawler's meta-search technology highlights and uses the strengths of major search engines to deliver more relevant and comprehensive results. The results returned from these search engines include commercial (sponsored) and non-commercial results. The

current WebCrawler product is designed to identify the intent of a user's search. If the search term is determined to be commercial in nature, a mixture of results with several sponsored results are displayed. In case of a non-commercial term, a mixture of results with several research and education material is displayed.

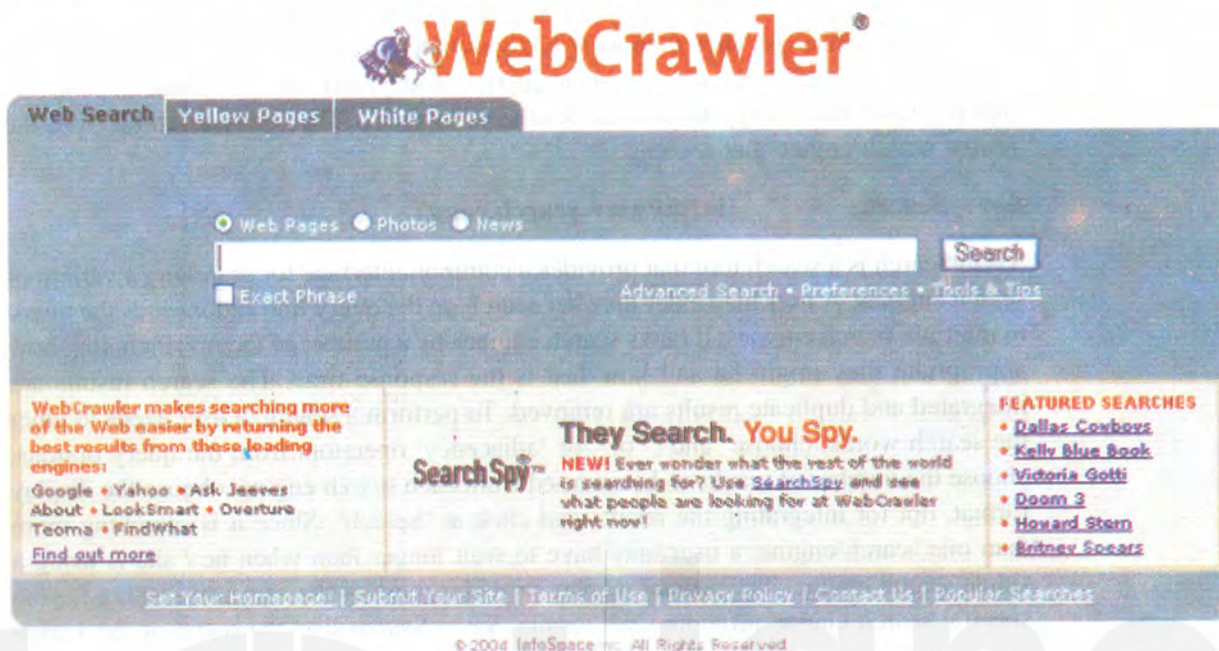


Fig. 13.21: WebCrawler Search Engine

HotBot (<http://www.hotbot.com/>)

HotBot, owned by Terra / Lycos, is one of the older web search engines. Originally, it just used the Inktomi database and then added Direct Hit and the Open Directory. In December 2002, it was re-launched as a multiple search engine with Inktomi, Fast, Google and Teoma. HotBot also uses other primary search engines to conduct searches. HotBot offers the choice of three search engine databases, namely HotBot (which is actually a Yahoo!/Inktomi database), Google and Ask Jeeves (the Teoma database). A user can switch between these databases with a simple click of button. Sponsored links on the top come from Overture while the side ones may come from other advertisers. While HotBot (Inktomi) includes some PDF, MS Word, PowerPoint, and Excel files, there is no limit for searching them. HotBot supports Boolean searching, proximity searching, field searching, limits to language, file types, page content, domain name, region, date, etc.

Results are sorted by relevance. However, only a limited number of results per domain are displayed, often two to four. The display includes the relevance score, title, URL, a brief extract, and date. HotBot displays 10 records at a time, by default. However, users can request display of 10, 25, 50, 75, or 100 records at a time. More search engines should give such options.

MetaCrawler (<http://www.metacrawler.com/>)

MetaCrawler is a search service that has no internal databases. It simply acts as a front end for 9 different search engines: OpenText, WebCrawler, Inktomi, Alta Vista, InfoSeek, Yahoo, Lycos, Excite and Galaxy. MetaCrawler sends the query from a user to the search engines, then puts them into a uniform format for display. The search screen offers a number of options. There is the usual search line but beneath it are 3 search options: search as a phrase, search all these words, search any of these words. Below these search options are options to limit by regions of the world, by type of site, by the maximum amount of time you want to wait for results and by the minimum score. The results display the title of the document, selected text or an abstract (depending on the search engine), the relevancy ranking, the URL, and the search engine from which the information came.

Ask Jeeves (<http://www.askjeeves.com>)

Ask Jeeves initially gained fame in 1998 and 1999 as being the 'natural language' search engine that facilitates search by asking questions and responds with what seemed to be the right answer to everything. In reality, technology was not what made Ask Jeeves perform so well. Behind the scenes, the company at one point had about 100 editors who monitored search logs. They then went out onto the web and located what seemed to be the best sites to match the most popular queries. Today, Ask Jeeves instead depends on crawler-based technology to provide results to its users. These results come from the Teoma search engine that it owns.

SavvySearch (<http://www.search.com/>)

SavvySearch is a search tool that provides a common interface for searching a variety of search engines. A user may enter his / her search on the query line and it sends the query to multiple search engines. It ranks search engines by a number of factors, including how appropriate they might be and how fast is the response time. The search results are integrated and duplicate results are removed. To perform a search, a user needs to enter the search words, choose 'and', 'or', or 'adjacency' operators from the query options, choose the number of results to be returned from each search engine, choose the display format, opt for integrating the results and click at 'Search'. Since it is searching more than one search engine, a user may have to wait longer than when he / she is using a single search engine. The normal display provides most of the standard display for the specific search engine providing the results. SavvySearch lists the name of the search engine providing the results. SavvySearch is currently available in 18 different languages.

13.11.4 Subject Portals or Subject Gateways

One of the most interesting, logical developments has been the rise of specialist index or gateway. These are directories of Internet resources in particular subject areas, compiled and organised by specialists and information professionals in their respective fields, which typically represent the best of sites available in the field concerned. Some of the important subject portals or subject gateways are given below:

LibrarySpot.com (<http://www.libraryspot.com/>)

LibrarySpot is a free virtual library resource centre for educators and students, librarians and their patrons, families, businesses and just about anyone exploring the web for information. LibrarySpot.com aims at breaking through the information overload of the web and bring the best library and reference sites together. Sites featured on LibrarySpot.com are hand-selected and reviewed by an editorial team for their exceptional quality, content and utility. Published by StartSpot Mediaworks, Inc. in the Northwestern University / Evanston Research Park, LibrarySpot is the first in a family of vertical information portals designed to make finding the best topical information on the Internet a quick, easy and enjoyable experience. The LibrarySpot.com has received more than 30 awards and honours. Most recently, Forbes.com selected LibrarySpot.com as a 'Forbes Favourite' site, the best in the reference category, and PC Magazine named it one of the top 100 websites. LibrarySpot.com has been featured on CNN, Good Morning America, CNBC and in many other media outlets.

Librarians' Index to the Internet (LII) (<http://lii.org/>)

The Librarians' Index to the Internet (LII) consists of more than 8,600 Internet resources selected and evaluated by librarians for their usefulness to users of public libraries. Free e-mail subscription to the LII New This Week (<http://www.lii.org/search/ntw>) incorporates most recent resources added to the LII. It has close to 12,000 subscribers in 85 countries. ILL also offers co-branding service to the libraries that are members of the Library of California. The site provides both browsing and searching interfaces.



Fig. 13.22: Librarian's Index to the Internet

Argus Clearing House (<http://www.clearinghouse.net/>)

The Argus Clearing House is a guide to the meta resources. It provides a central access point for value-added topical guides that identify, describe, and evaluate Internet-based information resources. The Argus Clearinghouse is a non-profit venture run by a small group of dedicated individuals. It is intended to be a resource that brings together finding aids for students, researchers, educators, and others interested in locating authoritative information on the Internet.

Vlib: The Virtual Library (<http://www.vlib.org/>)

The Virtual Library is the oldest catalogue of the web, started by Tim Berners-Lee, the creator of html and the web itself. Unlike commercial catalogues, it is run by a loose confederation of volunteers, who compile pages of key links for particular areas in which they are expert; even though it isn't the biggest index of the web. The Virtual Library pages are widely recognised as being amongst the highest-quality guides to particular sections of the web. Individual indexes live on hundreds of different servers around the world. A set of catalogue pages linking these pages is maintained at <http://vlib.org>. Mirrors of the catalogue are kept at East Anglia (UK), Geneva (Switzerland) and Argentina. Each maintainer is responsible for the content of their own pages, as long as they follow certain guidelines. The central affairs of the VL are now coordinated by a newly-elected Council.

Academic Info (<http://www.academicinfo.com/>)

Academic Info, online since 1998, began as an independent Internet subject directory owned by Michael Madin and maintained with the assistance of a quality group of subject specialists. In the spring of 2000 Michael left the University of Washington Gallagher Law Library to focus solely on Academic Info. In 2002 Academic Info became a registered non-profit organisation of the State of Washington. Academic Info is now ad-free and relies on donations to remain online. Academic Info aims to be the premier educational gateway to online high school, college and research level Internet resources. The primary focus of the site is academic, with its intended audience at the upper high school level or above. A priority is adding digital collections from libraries, museums, and academic organisations and sites offering unique online content. The current focus is on English language resources but selectively sites in other languages will be considered.

BUBL (<http://bubl.ac.uk/>)

BUBL LINK is the catalogue of selected Internet resources covering all academic subject areas and catalogued according to DDC (Dewey Decimal Classification). All items are selected, evaluated, catalogued and described. Links are checked and fixed each month. LINK stands for Libraries of Networked Knowledge. BUBL 5:15 provides an alternative interface to this catalogue, based on subject terms rather than DDC. The aim is to guarantee at least 5 relevant resources for every subject included, and a maximum of 15 resources for most subjects, hence the name 5:15. Big subject areas are broken down into smaller categories. However, the upper limit of 15 is not rigidly applied, so there may be up to 35 items for some subjects. The subject terms used in BUBL LINK / 5:15 were originally based on LCSH (Library of Congress Subject Headings) but have been heavily customised and expanded to suit the content of the service. The aim is to make it very easy to locate Internet information about a large number of subjects. The BUBL LINK catalogue currently holds over 11,000 resources. This is far smaller than the databases held by major search engines, but it can provide a more effective route to information for many subjects, across all disciplines.

BIOME (<http://biome.ac.uk/>)

BIOME is a collection of gateways, which provide access to evaluated, quality Internet resources in the health and life sciences, aimed at students, researchers, academics and practitioners. A core team of information specialists and subject experts based at the University of Nottingham Greenfield Medical Library creates BIOME. The Internet resources are selected for their quality and relevance to a particular target audience. They are then reviewed and resource descriptions created, which are stored, generally with the associated metadata, and generally in a structured database. The consequence of this effort is to improve the recall and especially the precision, of Internet searches for a particular group of users. BIOME is a hub within the Resource Discovery Network (RDN) (<http://www.rdn.ac.uk/>), and is funded by the Joint Information Systems Committee (JISC) (<http://www.jisc.ac.uk/>). There are five dedicated subject services (gateways) within BIOME, each covering a specific area within the health and life sciences. These gateways are AgriFor, VetGate, OMNI, Natural Selection and Bio Research.

Edinburgh Engineering Virtual Library (EEVL) (<http://www.eevl.ac.uk>)

Edinburgh Engineering Virtual Library (EEVL) is an award-winning free service, which provides quick and reliable access to the best engineering, mathematics, and computing information available on the Internet. It is created and run by a team of information specialists from a number of universities and institutions in the UK for students, staff and researchers in higher and further education, as well as anyone else working, studying or looking for information in Engineering, Mathematics and Computing. EEVL provides a central access point to networked engineering, mathematics and computing information. Resources being added to the catalogues are selected, catalogued, classified and subject-indexed by experts to ensure that only current, high-quality and useful resources are included. They include e-journals, databases, training materials, professional societies, university and college departments, research projects, bibliographic databases, software, information services and recruitment agencies.

EEVL, in addition to Internet Resource Catalogues, provides targeted engineering search engines - to UK engineering sites, to engineering e-journals, engineering newsgroups, and to specialised information services, such as the Recent Advances in Manufacturing (RAM) bibliographic database, and the Offshore Engineering Information Service. MathGate at EEVL is involved in the Secondary Homepages Project for UK Mathematics Departments. EEVL's scope is limited to the three subjects, and is therefore more focused than the big search engines. Searching EEVL will retrieve high quality resources, but because EEVL's resources are handpicked, the number of sources covered in it is not comparable to the number of Internet search engines.

The Social Science Information Gateway (SOSIG) is a freely available Internet service which aims to provide a trusted source of selected, high quality Internet information for students, academics, researchers and practitioners in the social sciences, business and law. It is part of the UK Resource Discovery Network. The SOSIG Internet Catalogue is an online database of high quality Internet resources. It offers users the chance to read descriptions of resources available over the Internet and to access those resources directly. The catalogue points to thousands of resources, and each one has been selected and described by a librarian or academician. The catalogue is browsable or searchable by subject area. Social Science Search Engine is a database of over 50,000 Social Science web pages. Whereas subject experts have selected the resources found in the SOSIG Internet Catalogue, those in the Social Science Search Engine have been collected by software called a 'harvester' (similar mechanisms may be referred to as 'robots' or 'web crawlers'). All the pages collected stem from the main Internet catalogue and this provides the equivalent of a social science search engine.

13.12 SUMMARY

Search engines are tools to search diverse and disorganised sources of information available on the Internet. There are several search, resource discovery and browsing tools that have been developed to support more efficient information retrieval. Search engines use automated programs variably called as spiders, robots, crawlers, wanderers and worms. Search engines are defined as tools for finding, classifying and storing information about various websites on the Internet.

This Unit traces evolution of search engines from Archie, developed in 1990, which can be considered as the first search engine that was used for indexing and searching files on FTP server. Archie, VERONICA and Jughead can be considered as workhorse used for searching the Internet before the launch of World Wide Web in 1993. The first robot, called World Wide Web Wanderer, was introduced to search the web in 1993 itself. In October 1993, an Archie-like search engine, called ALIWEB was created. By December of 1993, three full-fledged robot-fed search engines had surfaced on the web, i.e., JumpStation, the World Wide Web Worm, and the Repository-Based Software Engineering (RBSE) spider. The Excite search software was released by mid-1993. The year 1994 witnessed the launch of two important web directories, i.e., the EInet Galaxy and Yahoo!. The WebCrawler was launched on April 20, 1994. It was the first crawler that indexed entire pages of the web. WebCrawler opened the door for many other services to follow the suit. Within one year of its launch came Lycos, Infoseek, and OpenText. Lycos and Infoseek which were the next major developments that came about in July 1994. AltaVista was launched in December 1995, which brought many important features to the web searching. They were the first to allow natural language queries and advanced search techniques. The Looksmart directory commenced functioning in 1996. The Inktomi Corporation came about in May, 1996 with its search engine Hotbot. Ask Jeeves was launched in April 1997 followed by the Northern Light. The year 1998 witnessed the launch of Google, the most powerful search engine till date. Three other major search engines and directories were launched in the same year. They were: MSN search, Open Directory and Direct Hit. Disney released the Go Network in 1999. Fast released its search technology in the same year and that was considered the closest competitor to Google. In 2000, the Teoma search engine was released, which used clustering to organise sites by subject-specific popularity. In 2001, Ask Jeeves bought Teoma to replace the Direct Hit search engine. The evolution of search engines is checkered with companies in the business of web search technology acquiring other companies to strengthen their own position.

Describing functioning of a search engine, the Unit elaborates the following three components of a search engine:

- i) **The Robots:** Robots or spiders traverse the web using links that are embedded in the web pages to find information and build indexes of visited web pages. Besides indexing web pages, a robot / spider also validates links and finds new and updated information on websites;
- ii) **Databases:** A search engine builds a database of index information harvested by its robots / spiders from web pages. The index information or metadata includes URLs, titles, headers, words from title and text, first lines, abstracts and some times even the full-text; and
- iii) **User Interface or Agent:** The user interface or the agent is software that searches through the database consisting of the index of millions of pages recorded in the index to find matches to a search and rank them in order of relevance. The agent also displays the results of the search in convenient ways to the users.

Search engines have their own methods of organising information. While some search engines provide only a search interface to their databases, others are organised as directories wherein a user can browse through categories and subcategories which, in turn, contains links to relevant websites. The Unit divides search engines into the following five categories:

- i) **Primary search engines** deploy computer programs called web crawlers or spiders to traverse the web and scan websites for words, phrases, or the whole site so as to generate a database of web pages. Google and Alta Vista are examples of primary search engines;
- ii) **Meta search engines** pass queries on to many search engines and web directories and present summarised results to the users. Ask Jeeves, Dogpile, Infind, Metacrawler, Metafind and Metasearch are examples of meta search engines;
- iii) **Specialised search engines** are primary search engines that focus on a small or specialised segment of the Internet. Examples of specialised search engines are Direct Search, Beaucoup, Hoovers Online and Sirus;
- iv) A **Web directory** contains information that is organised into categories and subcategories or directories. Like a search engine, one can search a web directory for all entries that contain a particular set of keywords. Directories differ from search engines in the way they organise information. Yahoo, Dmoz.org and LookSmart are examples of web directory; and
- v) A **hybrid search engine** organises its contents like a web directory and provides a search interface as well for a user to opt for either of the two. It is extremely common for search engines to present both types of functionalities to the users. There are many directories that attempt to organise the Internet by subject. Moreover, there are many search engines that combine directory and keyword search capability.

Primary search engines can further be divided into the following five categories according to the methods their robots use for collecting information from the web for their databases:

- a) **Automated Robots:** The automated robots scan a large part of web wherever they are allowed.
- b) **Designated Robots:** The designated robots are programmed to scan only specific sites on the web, like ALIWEB or W3 catalogue, to collect and verify data rather than traverse the entire web. Sites using 'designated robots' provide for users to submit their websites to the search engine. On submission of a URL, the new URL gets added to the robot's queue of websites to visit on its next foray out onto the web. Even if a user does not submit a site, a designated robot may pick it up from other sites that provides links to other sites.
- c) **Breadth-oriented Search Engines:** Some robots concentrate more on top-level resources that tend to store larger subject-oriented index engines, such as JumpStation

II. A query conducted on “breadth oriented search engines” would result in a fewer retrievals with a higher percentage of those being index-level resources.

- d) **Depth-oriented Search Engines:** Search engines employing “depth-oriented robots” (like web crawler) follow links to deeper levels, often pulling out individual items located on a server’s indexes and following any links to other servers. Depth-oriented robots have the tendency to retrieve duplicates or false hits. The main drawback to this type of search engine is the tendency to catalogue too much information.

The Unit elaborates upon the criteria namely: ease of use, comprehensiveness, quality of content, control over the search, flexibility in searching, assessment of relevance and informative presentation of results for choosing a search engine. It also elaborates on search techniques that can be deployed to search subject directories and search engines. Subject directories are primarily designed to browse by clicking at topics, subtopics and finally at links to resources that are listed under these topics or subtopics. Search engines offer two types of interfaces to search their databases i.e., basic search and advanced search. In basic search, the user just enters a keyword sifting through any pull-down menus for additional options. However, options for advanced search differs from one search engine to another, but some of the common features include ability to search more than one word, to confine the search to a specified field and to exclude words that are not required in a search by the user. A user may also search proper names, phrases and words that are found with in certain proximity to other search terms. The Unit also elaborates upon the advanced search facilities offered by the primary search engines. It also deals with concept searching and natural language searching. A successful search leads to a list of best matches that are displayed by a search engine in an ordered fashion. The Unit describes as to how search results are displayed by the search engines in terms of number of records per page, customisation possibilities with regard to number of records displayed and items included in the displayed results (i.e., title, URL, description, etc.) The Unit also discusses the order in which results are displayed either using frequency of keywords used in the web page, positioning of keywords on the web page (title, description, body, etc.) and frequency of a document with which it is linked to other documents.

The Unit also elaborates on use of meta tags in the web pages and how search engines use them for indexing the website. It describes criteria that can be deployed to assess the success of a search engine. Lastly, the Unit describes important search engines in all the five categories described above.

13.13 ANSWERS TO SELF CHECK EXERCISES

- 1) Search engines are websites that facilitate users to search resources on the Internet. Search engines maintain a database of web pages, along with keywords to describe their content. The keywords are derived from each web page, using ‘keywords’ assigned in meta tag and the contents of the page. All search engines are not alike. Search engines can be divided into five categories, namely i) Primary search engines, which deploy computer programs called web crawler or spiders to traverse the web and scan websites for words, phrases, or the whole site so as to generate database of web pages. Google and Alta Vista are example of primary search engines; ii) Meta search engines, which pass queries on to many search engines and web directories and present summarise results for the users. Ask Jeeves, Dogpile, Infind, Metacrawler, Metafind and Metasearch are examples of meta search engines; iii) Specialised search engines, which are primary search engines that focus on small or specialized segments of the Internet. Examples of specialised search engines are Direct Search, Beaucoup, Hoovers Online and Sirius; iv) Web directories, which contain information that is organized into categories and subcategories or directories. Like a search engine one can search a web directory for all entries that contain a

particular set of keywords. Directories differ from search engines in the way they organise information. Yahoo, Dmoz.org and LookSmart are examples of web directory; and v) Hybrid search engines, which organize their contents like web directories and provide search interface as well for users to opt for either of the two. It is extremely common for search engines to present both types of functionalities to the users. There are many directories that attempt to organize the Internet by subject, and there are many search engines that combine directories and keyword search capability.

- 2) Before the advent of the World Wide Web, most information resources on the Internet were hosted on the FTP servers and Gopher sites. The Archie, developed in 1990 was the first search engine that was used for indexing and searching files on FTP server. It was actually an indexing spider that visited each anonymous FTP site, read all directories and file names, and then indexed them in one large database. VERONICA (Very Easy Rodent-Oriented Netwide Index to Computerized Archives) was developed at the University of Nevada to search all menu items on Gopher servers. The JugHead (Jonzy's Universal Gopher Hierarchy Excavation And Display) was a powerful Gopher search tool that searched a specified Gopher site. It searched directory titles only, not the text of resources that appeared on the Gopher submenus. It is thus true that Archie, VERONICA and Jughead as search tools can be considered as real workhorses for searchers on the Internet before the web's spectacular growth.
- 3) The World Wide Web Wanderer, launched in 1993, soon after advent of World Wide Web can be considered as the first search engine.
- 4) A spider or web crawler is a program that exhaustively surfs the Internet to collect web pages and returns them to another program for processing. All of the Internet search engine sites rely on spider robots to discover new websites and add them to their index. The spider program ensures that all of the links function correctly and reports dead links. Spider or robots can be categorised into four different categories depending upon methods for collecting information for their indexing databases. *Automated Robots* scan a large part of the web wherever they are allowed. *Designated Robots* are programmed to scan only specific sites on the web rather than traverse the entire web. Sites using "designated robots" provide for users to submit their websites to the search engine. *Breadth-oriented Search Engines* concentrate more on top-level resources that tend to store larger subject-oriented index engines such as JumpStation II. A query conducted on "breadth-oriented search engines" would result in a fewer retrievals with a higher percentage of those being meta resources or subject portals. *Depth-oriented Search Engines* follow links to deeper levels. They pull out individual items located on a server's indexes and follow links to other servers. Search engines that deploy "depth-oriented search engines" have a tendency to catalogue too much information.
- 5) A typical search engine has following three components:

The Robots: Robots or spiders traverse the web using links that are embedded in the web pages to find information and build indexes of visited web pages. Besides indexing web pages, a robot / spider also validates links and finds new and updated information on websites;

Databases: A search engine builds a database of indexing information harvested by its robots / spiders from web pages. The indexing information or metadata include URLs, titles, headers, words from titles and texts, first lines, abstracts and some times even the full-text

User Interface or Agent: The user interface or the agent is a software that searches through the database consisting of index of millions of pages recorded in the index to find matches to a search and ranks them in order of relevance. The agent also displays the results on the search in convenient ways to the users.

- 6) Primary search engines deploy spiders or robots to collect information from web pages. Spiders or robots can be divided into four categories according to the methods used by them for collecting information from the web for their databases. See answer to Question 4 for different categories of robots.
- 7) A subject portal can be defined as an organised and structured guide to Internet-based electronic information resources that are carefully selected after a predefined process of evaluation and filtration in a subject area or specialty. Subject portals are often independent web-sites or part of an institution's or library's website that serve as a guide to Internet resources considered appropriate for their target audiences. Meta search engines pass queries on to many search engines and web directories and present summarised results to the users. Unlike subject portals, meta search engines do not own a database of links to resources. Moreover, unlike subject portals, meta search engines do not confine to any specified subject.

13.14 KEYWORDS

- AltaVista** : A popular search engine with a large database on the web. Until 1998, this search engine provided the search facility for Yahoo.
- Archie** : Archie is a software tool for locating files on anonymous FTP sites. Archie is an indexing spider that visits each anonymous FTP site, reads all the directory and file names and then indexes them in one large index. A user can then query Archie which checks the query against the index.
- Ask Jeeves** : A unique search engine that combines a knowledge base of answers to roughly seven million commonly asked questions. This service is also in use at Altavista.
- Boolean Operators** : Boolean operators are used to combine search terms to construct more complex searches in a database. Using the Boolean operators AND, OR, and NOT can help to expand or narrow down a search.
- Case Sensitivity** : The ability of a search tool to distinguish between uppercase and lowercase letters. Most search engines are not case sensitive. Search engines that are case sensitive will strictly follow the search request and they would return documents containing the words in the case in which they were entered in the search expression.
- Concept Searching** : A feature that enables a search engine to find synonyms in its database. When a word or phrase is typed, the search engine automatically searches for that word or phrase, plus words or phrases that may mean the same thing. For example, if the word AIDS is searched, the search engine would also look for the word HIV.
- DirectHit** : A system that monitors user's selection of results from a search engine. It counts results that are clicked most, and time spent by the visitors at that site. These figures are used for improving relevancy. It is used by HotBot and also as a plug-in to Apple's new innovative Sherlock Search System.
- Excite** : Regarded as one of the more popular search engines. It searches top search engines and compile them to give best possible results.

FTP

: File Transfer Protocol, a common method of moving files between two Internet sites. FTP is a special way to login to another Internet site for the purposes of retrieving and / or sending files. There are many Internet sites that have established publicly accessible repositories of material that can be obtained using FTP, by logging in using the account name anonymous. So, these sites are called anonymous ftp servers.

Gopher

: Invented at the University of Minnesota in 1993 just before the web, Gopher was a widely successful method of making menus of material available over the Internet. Gopher was designed to be much easier to use than FTP, while still using a text-only interface. Gopher is a Client and Server style program, which requires that the user have a Gopher Client program. Although Gopher spread rapidly across the globe in only a couple of years, most Gopher sites have now been converted into websites.

Inktomi

: A database of websites used by many search engines like Hotbot, MSN etc. Yahoo also uses Inktomi when no site in its directory matches the keyword being searched for.

Jughead

: Jughead is a computer program that searches a particular Gopher site. Jughead searches directory titles only, not the text of resources appearing on the Gopher submenus. Jughead has disappeared with Gopher sites.

Listservs

: A free software program used for automating the maintenance and delivery of e-mail mailing lists. Listservs exist for many different topics. Some are open meaning anyone on a list can send a message to the whole list. Some are closed, meaning only certain people can post information to them.

LookSmart

: A website directory that has commercial search listings.

Lycos

: One of the large and popular search engines, Lycos appears to be moving towards becoming a directory and is using the Open Directory for some search results.

Meta Tags

: HTML tags that contain information about the page but are not displayed in the browser. Placed in the header section of an HTML document, meta tags may contain information such as title, description and keywords that can be used by search engines. Meta tags that can affect search engine rankings are keywords and description. The meta KEYWORDS tag is used to group a series of words that relate to a website. These tags can be used by search engines to classify pages for searches. The meta DESCRIPTION is used to describe the document. The meta description is displayed in search engine results.

Metasearch

: A search service that sends keywords to several search services at the same time. Results are usually shown together, with a note about which service provided the items.

Open Directory Project

: A directory project run by thousands of volunteer editors. In principle, this is a very exciting and powerful way to organize the web.

Phrase searching

: Phrase searching retrieves only documents with the exact combination of words and in the exact order. Most search engines allow phrase searching. Phrase searching is used to narrow down search results.

Proximity Searching : Searching for terms that occur within a specified number of words in the text. Proximity searching is not as stringent as phrase searching which requires that the term occur in the text exactly as entered and in the same order.

Relevancy Ranking : The process of displaying search results in an order based on an algorithm that considers the number of occurrences of the search term(s) in the documents, density of the term(s), number of inbound links to the document along with additional defined factors in order to display the most relevant documents first on the results pages and in decreasing order of relevance down the list.

Robot : Any browser program which follows hypertext links and accesses web pages but is not directly under human control. Examples are the search engine spiders, the 'harvesting' programs which extract e-mail addresses and other data from web pages and various intelligent web searching programs.

Spider : A software program that 'crawls' the web, searching and indexing web pages to create a database that can be easily searched by a search engine

Truncation : A search technique in which a user adds a symbol (such as * or ?) to a word stem in order to retrieve multiple endings of the word. It is also called stemming. For example, child* will search for child and also children, child's, childless and any other words with child as the first five characters.

VERONICA : Stands for Very Easy Rodent-Oriented Net-wide Index to Computerized Archives. Developed at the University of Nevada, Veronica was a database of the names of almost every menu item on thousands of Gopher servers. VERONICA has disappeared since most Gopher sites have been converted into websites.

Wildcards : Wildcards is a symbol that facilitates search for multiple endings of a word. Wildcards can be used for external truncation or stemming or for internal truncation. External truncation is used to search for all possible endings to a word. Internal truncation is used to search for words that may differ by one or two letters within a word.

Yahoo : Acronym for "Yet Another Hierarchical organized Oracle". Yahoo is a web directory that categorizes web pages. It is one of the Internet's leading search tools.

13.15 REFERENCES AND FURTHER READING

Animated Internet: How Search Engines Work. (<http://www.learnthenet.com/english/animate/search.html>)

Balas, J. (1996). Beyond Veronica and Yahoo! More Internet Search Tools. *Computers in Libraries*, 16(3), 34-35.

Barlow, Linda. *The Spider's Apprentice: A Helpful Guide to Search Engines*. March 2, 2001. (<http://www.monash.com/spidap.html>)

Bates, M.E. (1997). The Internet: Part of a Professional Searcher's Toolkit. *Online*. 21(1).

- Brandt, D.S. (1996). Relevancy and Searching the Internet. *Computers and Libraries*. 16(8).
- Cohen, Laura. *Searching the Internet: Recommended Sites and Search Techniques*. May 23, 2001. (<http://library.albany.edu/internet/search.html>)
- Cohen, Laura. *Conducting Research on the Internet*. May 2001. (<http://library.albany.edu/internet/research.html>)
- Cohen, Laura. *Quick Reference Guide to Search Engine Syntax*. May 2, 2001. (<http://library.albany.edu/internet/syntax.html>)
- Cohen, Laura. *Second Generation Searching on the Web*. June 12, 2001. (<http://library.albany.edu/internet/second.html>)
- Dong, L.T. and Su, L.T. (1997). Search Engines on the World Wide Web and Information Retrieval from the Internet: A Review and Evaluation. *Online and CD ROM Reviews*, 21(2).
- Falk, H. World Wide Web and Retrieval. *Electronic Library*, 15(1), 1997.
- Grossan, Bruce. *What They are, How They Work, and Practical Suggestions for Getting the Most Out of them*. (<http://webreference.com/content/search/refer.html>)
- Hill, Brad (2003). *Google for Dummies*. New York: John Wiley & Sons, Inc..
- How to Use Web Search Engines* (<http://www.monash.com/spidap4.html>).
- Kent, Peter (2004). *Search Engine Optimisation for Dummies*. New York: John Wiley & Sons, Inc.
- Levine, John R: et.al. (2000). *The Internet for Dummies: Starter Kit*. 7th ed. New York: John Wiley & Sons, Inc.
- Liu, Jian. *Guide to Meta Search Engines*. (<http://www.indiana.edu/~librcsd/search/meta.html>)
- Lowe, Doug (2001). *Internet Explorer 6 for Dummies*. New York: John Wiley & Sons, Inc.
- Notess, Greg R. Search Engine Statistics: *Relative Size Showdown: Data from Search Engine Analysis*. Run on Dec. 31, 2002 (<http://searchengineshowdown.com/stats/size.shtml>)
- Search Engine Show Down* (<http://searchengineshowdown.com/>)
- Search Engines.com. *Search Engine Rankings and Search Engine Optimization Tips*. (www.searchengines.com/)
- Vine, Rita. Real People Don't Do Boolean: How to Teach End Users to Find High-Quality Information on the Internet. *Information Outlook*. March, 2001. (http://www.findarticles.com/p/articles/mi_m0FWE/is_3_5/ai_71965499)
- Tyner, Ross. *Sink or Swim: Internet Search Tools and Techniques*. 2001 (<http://www.sci.ouc.bc.ca/libr/>)

UNIT 14 INTERNET SERVICES

Structure

- 14.0 Objectives
- 14.1 Introduction
- 14.2 World Wide Web
 - 14.2.1 Importance of the Web
- 14.3 How does the Web Work?
 - 14.3.1 Client-Server Architecture
 - 14.3.2 Hypertext Transfer Protocol (HTTP)
 - 14.3.3 Hypertext Links: Uniform Resource Locators (URL)
- 14.4 Web Servers
- 14.5 Web Browsers
 - 14.5.1 Plug-ins or Helper Programs
 - 14.5.2 Using Web Browser
 - 14.5.3 Toolbar
 - 14.5.4 The Location (URL) Box
- 14.6 Mark-up Languages
 - 14.6.1 Standard Generalized Mark-up Language (SGML)
 - 14.6.2 Extensible Mark-up Language (XML)
 - 14.6.3 Hypertext Mark-up Language (HTML)
 - 14.6.4 Dynamic HTML
 - 14.6.5 Virtual Reality Modelling Language (VRML)
- 14.7 Internet Applications
 - 14.7.1 Internet-based Communication Services
 - 14.7.2 Connectivity
 - 14.7.3 Access to Information Resources
 - 14.7.4 Searching Information Resources on Internet
- 14.8 Internet for Library Applications
 - 14.8.1 Use of Internet for Supporting Traditional Library Activities
 - 14.8.2 Traditional Library Services Modified in the Internet Era
 - 14.8.3 Internet-based New Library Services
- 14.9 Summary
- 14.10 Answers to Self Check Exercises
- 14.11 Keywords
- 14.12 References and Further Reading

14.0 OBJECTIVES

After reading this Unit, you will be able to acquire knowledge on the following components of Internet:

- World Wide Web, its importance and its functioning;
- web servers and web browsers and using web browser;
- mark-up languages: SGML, XML, HTML, Dynamic HTML and VRML; and
- general Internet applications and library-specific Internet applications.

14.1 INTRODUCTION

The Internet and web hold tremendous potential for a whole array of activities including online distance education, global digital library, e-commerce, Internet telephony, electronic publishing, electronic journals, virtual museums, etc. It has particularly established itself as a powerful medium for self-education for people in isolated or remote areas for its ease-to-use, familiarity with masses, availability of tools and wider accessibility. The web has become the most successful networked multimedia hyper-text-based system of our time. HTML, the *de facto* language of the web, is extremely simple yet powerful to use. Further, the static HTML web pages can be transformed into vibrant, dynamic and interactive web creations using ever evolving web technologies like CGI Script, Perl, Java, Javascript, ASP, DHTML, XML and open database connectivity (ODBC) drivers.

The rapid growth in the web technology and its ever increasing usage has given librarians and educators with unprecedented opportunities to provide information to the students not within the four-walls of libraries and class rooms but also in the comforts of their home all over the globe. The changes, mainly driven by the new technological innovations and the new learning environment, has presented a scenario where students have access to a vast array of information in many fields from experts all over the world.

The Internet allows us to share information and resources such as, government documents, electronic journals, electronic books, media publishing, human anatomical images, computer software, bibliographic and full-text databases, speeches, live concerts, audio and video clippings. The dynamic nature of Internet is derived from scientists, other researchers and general public 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 publications, make travel arrangements and purchase goods and services. Electronic mail and news groups assist users to communicate with each other on topics of mutual interest. The discussion forum and listserv provides a platform to people with common interest to engage in thoughtful discussions. A few popular modes of usage of Internet are as follows:

- Retrieving information from reference sources like dictionaries, encyclopaedias, etc. available on the web, for use in day-to-day work;
- Retrieving information from databases of various libraries like the Library of Congress, the British Library, and those of Indian Institute of Science, IITs and several other institutions;
- Searching commercial and non-commercial databases like MEDLINE, INSPEC, COMPENDEX, etc.;
- Accessing electronic books, e-journals and other e-documents available on the websites of commercial and non-commercial publishers, as required for research work;
- Referring social, economic and statistical data, such as, census, 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 people across the world;
- Searching for computer shareware, freeware, and commercial software;
- Sending or receiving sound, animation and picture files across the Internet;
- Setting up temporary or permanent discussion or work-oriented groups;

- Distributing or reading electronic newsletters, newspapers, bulletins and other publications and marketing products and services;
- Trading with people of other organisations and other e-commerce activities; and
- Chatting with people using software like Yahoo Messengers, Hotmail Messengers, etc.

The explosive growth of the Internet and the World Wide Web in recent years has its impact on the information profession too. It has registered a sea change in the information seeking approach as well as the mode of dissemination of information. As librarians and information professionals, our prime responsibility is to acquire, organise, preserve, retrieve, and disseminate pertinent information to our clientele. This global forum, an emerging medium of communication, and a proven and concrete technology in sharing and exchanging information, has a lot to offer to the information professionals.

The Internet works on client-server technology, i.e., it works on two types of computer programs, i.e., servers and clients. Servers are programs that host resources to serve the clients and clients are programs that users use to access these resources. E-mail, listserv / mail lists, Usenet / newsgroups, FTP, Telnet, Gopher, Archie, WWW, etc., are among the prominent services of the Internet. Each type of service in the Internet has its own client. For example, to access the WWW, we need to use a web client such as, 'Netscape' or 'Internet Explorer'. The Internet is a network of networks that connects thousand of networks all over the world. Different types of computers on these networks are made to work seamlessly using TCP/IP (Transmission Control Protocol / Internet Protocol). The TCP / IP protocol, in turn, is the common name for a collection of more than 100 protocols used to connect computers and networks.

14.2 WORLD WIDE WEB

The World Wide Web, known as WWW, W3 or simply, the web, is one of the several Internet resources developed to help people publish, organise and provide access to information on the Internet. The web was first developed by Tim Berners Lee in 1989 while working at CERN, European Particle Physics Laboratory in Switzerland, and has since become the most powerful, and popular, resource discovery tool 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.

14.2.1 Importance of the Web

The World Wide Web is important for libraries because it provides an extremely powerful method of organising and providing access to information. The Web can provide a single interface to a large variety of information resources and systems including textual (unformatted or formatted) documents, images, sound and video files. The web can be used to provide interface to other Internet services like TELNET, FTP and Gopher. It can also be interfaced to online databases. There are several features unique to the web that makes it the most advanced hypertext-based information system on the Internet. These features are briefly discussed below:

The Web is a Hypertext System: The web is a hypertext system, in contrast to the hierarchical menu system used by earlier Internet tools such as, Gopher. The user on the web moves from one document to another related document through embedded links (called hyperlinks) and a hyperlinked word or phrase, when clicked, calls for 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 clicking on hypertext links.

The Web is a Multimedia System: The web is the most successful networked, multimedia, hypertext-based system of our times. The web technology allows incorporation of various

media types besides structured text. A good multimedia interactive document is a product consisting of structured text, video clips, animation, pictures, graphics, diagrams, programs, sound, etc. With the advent of graphical browsers, the web has become a multimedia system, combining different types of media into one document. Before the advent of graphical web browsers (e.g., Netscape, Internet explorer), most of the information available on the Internet was in the form of simple text devoid of any elements common to the printed page, such as, text in bold and italics, pictures and other graphical contents. The web documents may contain the following:

- Normal text;
- Features such as, large fonts, bold, italics, indents;
- Images such as, pictures, graphics, logos, illustrations;
- Audio content such as, sounds, music, commentary, voice messages; and
- Video content such as, movie clips, animations, or computer-generated simulations;

The Web is a Distributed System: The web is a distributed system for delivering linked documents over the Internet. It is called a distributed system because information can reside on different computers around the world, yet it can easily be linked together using hypertext. The web uses hypertext to create links from one resource to another. From the perspective of users, 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.

The Web incorporates other Internet Tools: The web incorporated the capabilities of most of the earlier tools, and added the ability to handle various media types. The web can provide links to other types of Internet tools, such as, WAIS, Gopher, FTP and TELNET. A web page can provide links 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.

The web provides an Interface to other database systems: A particularly powerful feature of the web is that it can act as an interface to database systems connected to the Internet. There are three elements that are needed to create this interface.

- i) Forms: Forms are used to collect information through web browser. Forms are a method of creating input boxes on a web page into which users can type information, or select among alternatives.
- ii) Database System: RDBMS system such as, MS Access, MS SQL, MySQL, Oracle or PostGres can be used as back-end database.
- iii) Control Gateway Interface (CGI): The CGI sits between the web browser and database. It takes the information gathered from the web browser and passes it to the database. Once the request is processed, the CGI passes the result back to the web browser in a format that it can display.

Computer and communication technology with its capabilities of parallel processing, multitasking, parallel consultation and parallel knowledge navigation, put together, creates a semblance of artificial intelligence and interactiveness necessary for developing an interactive learning interface. Coincided with availability of software, hardware and networking technology, the advent of World Wide Web, its ever increasing usage and highly evolved browsers have paved the way for creation of a global digital library. The increasing popularity of the Internet and developments in web technologies act as catalyst to the development of highly interactive library services.

1) What is World WideWeb? Why is it called multimedia hypertext system?

Note : i) Write your answer in the space given below.

ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

14.3 HOW DOES THE WEB WORK?

The most important concepts and underlying mechanism that make the web work are client-server architecture, the Hypertext Transfer Protocol (HTTP), Hypertext Mark-up Language (HTML) and Universal Resource Locators (URLs). These concepts are described below:

14.3.1 Client-Server Architecture

The Client-Server Architecture is based on the principle where the 'client' program installed on the user's computer (called client) communicates with the 'server' program installed on the host computer to exchange information through the network. The client-server model involves two separate but related programs, i.e., client and server. The client program is loaded on the PCs of users hooked to the Internet where as the server program is loaded on to the 'host' (usually a PC with large storage capacity and RAM, a mini-computer or a main-frame computer) that may be located at a remote place. The concept of client / server computing has particular importance on the Internet because most of the programs are built using this design. The client server architecture is discussed in detail in Block 4, Unit 12 of this course.

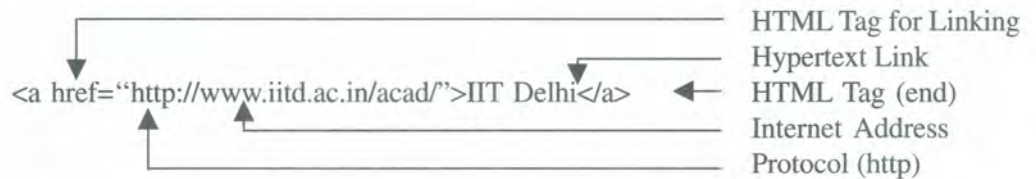
14.3.2 Hypertext Transfer Protocol (HTTP)

The Hypertext Transfer Protocol (HTTP) is a set of rules for exchanging files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web. As its name implies, the essential concept of HTTP is the idea that files can contain links or references to other files whose selection would lead to transfer of requests from one file to another. Any web server machine contains, in addition to the HTML and other files it can serve, an HTTP daemon, a program that is designed to wait for HTTP requests and handle them when they arrive. The web browser is an HTTP client, sending requests to server machines. When a user requests for a file through browser by either 'opening' a web file (typing in a Uniform Resource Locator) or clicking on a hypertext link, the browser builds an HTTP request and sends it to the Internet Protocol address indicated by the URL. The HTTP daemon in the destination server machine receives the request and, after any necessary processing, the requested file is returned. HTTP protocol is Please refer to Block 4, Unit 12 of this course.

14.3.3 Hypertext Links: Uniform Resource Locators (URL)

Hypertext links are words, phrases, symbols, maps or any other item in a web document that are linked to a different place in the same document or to another Internet resource. Hypertext links may be underlined, highlighted in colour, or appear as icons, to distinguish them from the surrounding text. The link must be 'selected' by clicking on it with a mouse so as to call upon another document or part of a document. The

hypertext inks embed an URL into an object (such as, text or an image). The URL is a compact string representation for a resource available on the Internet. 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 link to the URL is:



The link shown above has four parts:

- i) The protocol used to connect to the remote server. In this example, the protocol is HTTP; the protocol is 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;
- ii) The Internet address of the server where the document resides. In this case, the address is <http://www.iitd.ac.in>;
- iii) The directory on the server where the document is located, called the document path. In this case, the path is /acad/; and
- iv) The filename of the document itself. In the example, it is index.html (default file) where the html extension indicates that the document is marked-up with HTML tags.

14.4 WEB SERVERS

The Internet works on client-server model. A server is a computer system that is accessed by other computers and / or workstations at remote locations. Usually, a server contains data, datasets, databases and programs. The server computer is also called 'host' since it is configured to host datasets, files and databases, receive requests for it from the client machines and serve them. The term 'host' means any computer that has full two-way access to other computers on the Internet. All computers that host websites are host computers or servers since they 'host' information and 'serve' client machines. Clients and servers are two ends of the web, each with its own supporting software. A web server is a software application that uses the Hyper Text Transfer Protocol. There are many web server software applications, including public domain software from Apache, and commercial applications from Microsoft, Oracle, Netscape and others. A web server may host or provide access to content and respond to requests received from web browsers. Every web server has an IP address and usually a domain name, e.g. www.iitd.ac.in. 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-user's computer accessing, translating and displaying documents.

Web servers process HTML documents for viewing by web browsers. The server enables users on other sites to access documents that it hosts. web servers can run from any hardware platform. There are servers that are specifically designed for Macintosh computers, PCs, Silicon Graphics, and various other platforms. The most important software is the web server itself. Just like a web server can run on a number of hardware platforms, it can also run under several operating systems, including MS Windows, Windows NT, Unix, Linux and Macintosh.

A 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 communicate

HTTP in order to send and receive hypermedia documents. For this reason, web servers are often called HTTP servers, or HTTP Deamons (HTTPD).

Web documents are written in a text-formatting language called Hypertext Mark-up Language (HTML). The HTML is used to create hypertext documents for use on the web. Basically it is a set of 'mark-up' symbols or codes inserted in a web file that tells the web browser how to display a web page for the user. HTML is a language, neither an application nor a software package. It is simply a data set of text and instructions that requires a web browser to be used.

Self Check Exercise

- 2) What is Client- Server architecture?
- 3) Describe HTTP protocol. How does it work?
- 4) What is URL?

Note : i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

14.5 WEB BROWSERS

Web browsers are the applications that allow a user to view HTML documents from a computer connected to the Internet. Software such as, Netscape, Microsoft Explorer, etc., read files created with HTML (Hyper Text Mark-up Language) and display interactive web pages to the user. The first browser, called NCSA Mosaic, was developed at the National Center for Supercomputing in the USA. The easy-to-use point-and-click interface helped to popularise the web. Availability of ready-to-use, publicly available, user-friendly graphical web browsers for all prevalent platforms paved the way for unprecedented growth of Internet applications and services. Standard WWW clients such as, Netscape Navigator and Internet Explorer are being upgraded regularly for added functionality such as, e-mail client, support for JAVA and Active X and the ability to view important document formats without having to install plug-ins for them. These browsers solved the maintenance problem allowing developers to concentrate fully on the server side and not to bother with the client side. These browsers are available freely and are easy to use eliminating the need of extensive support and user training. The two important graphical browsers are Microsoft Internet Explorer and Netscape Navigator. Both are fast and both have integrated audio and video. Most of the browser can be downloaded at no charge.

There is no standard way of viewing or navigating the web. A variety of web browsers exist. Most browsers have most of the functionality, although there are some differences in levels of support and overall performance. Most browsers are still being updated and improved, with new releases every two or three months. A number of web browsers are available for each computing platform including Lynx for terminal-based users (without the graphics support). 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.

A web browser is a client program that uses the Hypertext Transfer Protocol (HTTP) to make requests to the web servers on behalf of the user. A commercial version of the original browser, Mosaic was launched as Netscape Navigator. Many of the user interface features in Mosaic, however, went into the first widely used browser, Netscape Navigator. Microsoft followed with its Internet Explorer. Today, these two browsers are highly competitive and most of the Internet users are aware of these two browsers only. Although the online services, such as, America Online, Compuserve and Prodigy, originally had their own browsers, virtually all now offer the Netscape or Microsoft browser. Lynx is a text-only browser for UNIX shell and VMS users. Another recently offered browser is Opera.

A web browser contains the basic software a user needs in order to find, retrieve, view, and send information over the Internet. Some of the important functions of a browser are:

- Send and receive electronic mail;
- Read messages from newsgroups, forums about thousands of topics in which users share information and opinions;
- Browse the World Wide Web (or web) where a user can find and view a rich variety of text, graphics and interactive information;
- Browsers such as, Microsoft Internet Explorer version 5.0 + include additional Internet-related software. For example, with Internet Explorer 5.0, also incorporates:
 - Windows Media server
 - NetMeeting, conferencing software
 - ActiveX.controls
 - Chat
 - ActiveMovie application programming interface
 - Active Channel webcast
 - Subscriptions
 - Dynamic HTML
 - Windows Media

The features mentioned above allow a user to see and hear live and recorded broadcasts such as, concerts or breaking news with synchronised audio, graphics, video, URLs, and script commands. Streaming technology allows a user to see or hear the information as it arrives instead of having to wait for the entire file to download.

The browser performs two tasks: first it identifies HTML elements, and then it translates the identified elements into actions. For example, it may identify the HTML bold element and then display a block of text in bold format. Other actions might be to display an image, add a blank line between text, or link to another document. Many of these actions can be handled by the browser itself; for example, within its viewing area it can display text and some types of images (if it uses a graphical user interface such as, Windows).

In addition to web servers, web browsers can also access Gopher, FTP, and WAIS servers. Essentially, besides, HTTP browsers understand protocols associated with Gopher, FTP and WAIS. Thus, browsers provide a common navigational interface between all these systems, seamlessly executing the appropriate protocol behind the scenes.

There are several web browsers or web clients available to surf the Internet.

Some of the important ones are:

- **Mosaic Version 2.1.1** (<http://archive.ncsa.uiuc.edu/SDG/Software/WinMosaic/HomePage.html>)

At one time first and foremost among web clients was the Mosaic graphical interface. It was developed in 1993 by the National Centre for Supercomputing Applications (NCSA) at the University of Illinois. Before Mosaic, all interfaces to the web were simple text-based, line-by-line interfaces. They were hypertext, but not graphical or multimedia. When the Windows version of Mosaic became available to Internet users for free, suddenly the web became the hottest information system on the Network because it was so much more powerful.

- **Netscape Navigator 7.1** (<http://www.netscape.com/>)

The Netscape Navigator was developed by the same people who created Mosaic at NCSA. The Netscape 7 browser has a tabbed user interface that allows easy switching from one open web page to another. A user can create bookmarks that open a specific set of tabs. Another convenient feature is one-click search. Highlight a word (but not a link) in the browser window and right-click on it, and start a search for it from any search engine. The new version of Netscape has also implemented the 'Sidebar' pane that runs down the left edge of the screen. Internet Explorer does the same thing (click on Favourites or History and a thin left-hand window opens). Netscape's Sidebar provides tabbed access to addresses, bookmarks, news, history, and a variety of other useful things. The Netscape has a mail client and address book that is quite adequate. The biggest difference between Netscape 7 and Internet Explorer is integration. Because Microsoft presumes ownership of your desktop, it has less need to pack as many applications into its web browser. Netscape, on the other hand, ties together e-mail, browsing, and instant messaging in one application. For example, both AOL's instant messaging client and ICQ are integrated into the browser. Microsoft's Messenger client is not fully integrated. In Netscape 7, the integration feels smooth and natural.

- **Internet Explorer 6.0** (<http://www.microsoft.com/windows/ie/>)

The Internet Explorer is Microsoft's Internet browser that comes packaged with Windows Operating System. It can also be downloaded from their website free of cost. Over 75% of Internet users use the Internet Explorer. The new version of Internet Explorer integrates Outlook Express. It provides a private, reliable, and flexible browsing experience and the freedom to experience the best of the Internet for users of Windows XP, Windows Millennium Edition (Windows Me), Windows 2000 Professional, Windows 98 and Windows 98 Second Edition, and Windows NT 4.0 workstation. It also includes a free copy of the Advanced Searchbar, which is an Internet Explorer toolbar that allows users to quickly access and search over 60 search engines and is jam packed with features including blocking of pop-up windows.

- **Avant Browser v9.02** (<http://www.avantbrowser.com/>)

This browser add-on, runs on top of Internet Explorer. An integrated pop-up stopper and Flash animation filter protect users from unwanted distractions. Avant Browser supports tabbed-multi-window browsing, i.e., the tabbed interface lets a user open several sites inside one browser and makes navigation easier. The built-in Google search engine lets a user search right from the browser's taskbar. It has several built-in features like records eraser to keep privacy by deleting typed addresses, auto-complete passwords, cookies, history of visited websites, temporary Internet files and search keywords. It has built-in Flash Animation filter and options to block downloads of pictures, videos, sounds and ActiveX components. With these options users can control their bandwidth and speed up page loading. Additional mouse functions such as, if link is clicked with the middle mouse button, the link will be opened in a new window in the background. It supports Real Full Screen Mode and Alternative Full Desktop Mode. It is fully Internet Explorer compatible and supports all Internet Explorer functions, including Cookies, ActiveX Controls, Java Script, Real player and Macromedia Flash. Internet Explorer's favourites are automatically imported to Avant. Avant Browser supports many different languages.

- **Enigma Browser** (<http://www.suttondesigns.com/>)

Enigma Browser incorporates a large collection of powerful features like built-in pop-up stopper, skinned window frame, form filler, site group, quick-search, auto login, hidden sites, built-in commands and scripting, online translation, script error suppression, blacklist / white list filtering, URL Alias. It provides convenient and comfortable browsing. It has ability to turn on / off Flash Animation. Enigma provides convenient access to major search engines by Quick-Search bar. It has built-in VBScript, Jscript, HTML and Text Editor. It has features to hide sites and show a site at user's request. It provides for auto login, i.e., it automatically connects and log into specified website with just one click. Enigma seamlessly integrates with online translation engine and dictionaries. It has the ability to suppress script error message dialogue.

- **Crazy Browser v1.05** (<http://www.crazybrowser.com>)

Crazy Browser facilitates browsing multiple websites at once. It blocks advertisements. Users have the option to turn off multimedia and browse the web in text mode. Users can search on a number of search engines that come with the program. It incorporates Smart Pop-up Filter. It supports tabbed-multi-window browsing.

- **AutomaticSearch Browser** (<http://www.4comtech.com/>)

AutomaticSearch is a search-themed web browser that automatically finds related links, subjects, and topics associated with the current website being viewed. It features an integrated search engine utilising the popular engines (Dmoz, Google, Yahoo, All the web, MSN, Lycos, Hotbot, etc.) and allows users to quickly switch among the search results using tabs. Users can also save and access their favourite websites easily by using toolbar buttons.

- **Mozilla v1.7.2** (<http://www.mozilla.org>)

Mozilla, developed by the Mozilla.org open-source community, is a cross-platform product with support for Windows, Linux, and Macintosh 8/9/X. It incorporates a filter to stop pop-up advertisements. It supports tabbed multi-window browsing that lets a user open new pages easily instead of forcing a user to open a new window and then click back to see the previous screen while the new window loads. Mozilla offers a tab to the new window and loads that page in the background, letting you stay focused on the work at hand. Another welcome advance is the ability to turn off pop-up ads and animated GIFs. Besides the browser, Mozilla includes an instant-messaging application, an e-mail client, web-composer software, and some nifty cookie-management and anti-spam features.

- **Opera** (<http://www.opera.com/>)

Opera browser is designed for Windows with built-in pop-up stopper to stop unwanted pop-up pages. It supports tabbed multi-window browsing. It has integrated Google search engine that lets a user search right from the browser's taskbar. It has a built-in mail client. Opera is extremely well designed with minimum resource use in terms of hard disc and memory requirement. Opera is fully customisable and it has page magnification capability, and graphics handling which make Opera an alternative to Internet Explorer or Netscape.

- **EasyBrowse v2.0** (<http://www.vrameen.com/>)

EasyBrowse is a free Internet browser that works on all Windows platforms (95 and above). The new version includes the Pop-up disabler option to a standard list of features including plug-ins, direct linking to 11 search engines, plug-in checker, large browsing window, favourites and a history list.

- **NeoPlanet Browser** (<http://www.neoplanet.com/>)

NeoPlanet offers a great browsing experience while allowing users to customise the interface. The NeoPlanet browser has an embedded e-mail client that has most of what a

user needs. Neoplanet facilitates download management and offers QuickSearch features. A key difference between NeoPlanet and other alternative browser options like Opera is that NeoPlanet is not actually a full-fledged web browser. NeoPlanet is actually a front-end for the Internet Explorer. It means that every technology supported by Internet Explorer is supported by NeoPlanet as well, i.e., from Java applets to dynamic HTML to ActiveX Controls. NeoPlanet also offers a powerful search function and a customisable channel bar to deliver “the best of the web in 3 clicks or less”.

- **Lynx** (<http://www.browser.lynx.org/>)

Lynx is a text-based, full screen interface to the web. Arrow keys, tabs, and the cursor are used to move around and select items instead of a mouse. Lynx interface is not multimedia, and so pictures, icons, maps and other graphical elements cannot be viewed.

14.5.1 Plug-ins or Helper Programs

The web is practically an ultimate tool of integration. It facilitates incorporation of all kinds of media files on to a web page. A web author can incorporate text files (formatted or unformatted), images, video clippings, audio files, graphics, animations, and other types of actions in a web page. The browser cannot handle all these files and formats and thus requires additional software programs to execute them.

Initially, the Netscape browser allowed users to download, install, and define supplementary programs that played sound or motion video or performed other functions. These were called helper applications. However, these applications run as a separate application and require that a second window be opened. A plug-in application is recognised automatically by the browser and its function is integrated into the main HTML file that is being presented.

Plug-ins or helper applications are external software programs that allow web users to view or hear multimedia presentations, regardless of platform. Plug-ins can easily be installed and used as part of web browser. Plug-ins or 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 over data to appropriate helper applications such as, RealAudio, Adobe Acrobat, QuickTime, Shockwave and others.

Native Helper Programs: Native helper programs are integrated into the browser itself, just like some word processors include internal spell check programs. In practice, the browser identifies an element, and then calls up a native helper program to execute an action. For example, when a browser identifies a file stored in the JPEG format (a type of compressed image), it calls up an internal program that can translate JPEG files. The internal program then processes and displays the image inside the browser's viewing area. Netscape has a native helper program that can handle JPEG files this way. Browsers that do not have the same capacity must use external helper programs to read images.

External Helper Programs: External helper programs address the fact that there are too many file formats for one browser to handle alone. Instead of one massive omnilingual program that can read every type of file the WWW carries, browsers incorporate smaller external helper programs to accomplish the same end. These specialised programs are separate from the browser and perform functions identical to native helper programs, except that the actions are executed outside of the browser. For example, when a browser identifies a sound file stored in .WAV format, it calls up an external program that can translate .WAV files. The browser passes the .WAV file to the external helper program, which then processes and plays it.

There are two main differences between native and external helper programs. The first difference is that external helper programs run independently of the browser. This means that once files are passed from the browser to the external helper program, the browser is free to resume navigating the WWW. In contrast, native helper programs tie up the

browser until the native program completes its action and is closed. The second difference concerns how the two types of helper programs are acquired. Native helper programs are included within the browser itself. However, external helper programs must be acquired independently by the end-user. The end-user then has to configure the browser to point to the external program, telling it when to use it (e.g., to view a particular file format) and where it is located within the computer's storage. Most often this is done through setting the Preferences area of the browser's Options.

14.5.2 Using Web Browser

A user should know how to use and move around the web with the web browser. It can really make using the web much easier. Although the specific features might differ from browser to browser, there are a few things that are common to virtually every web browser.

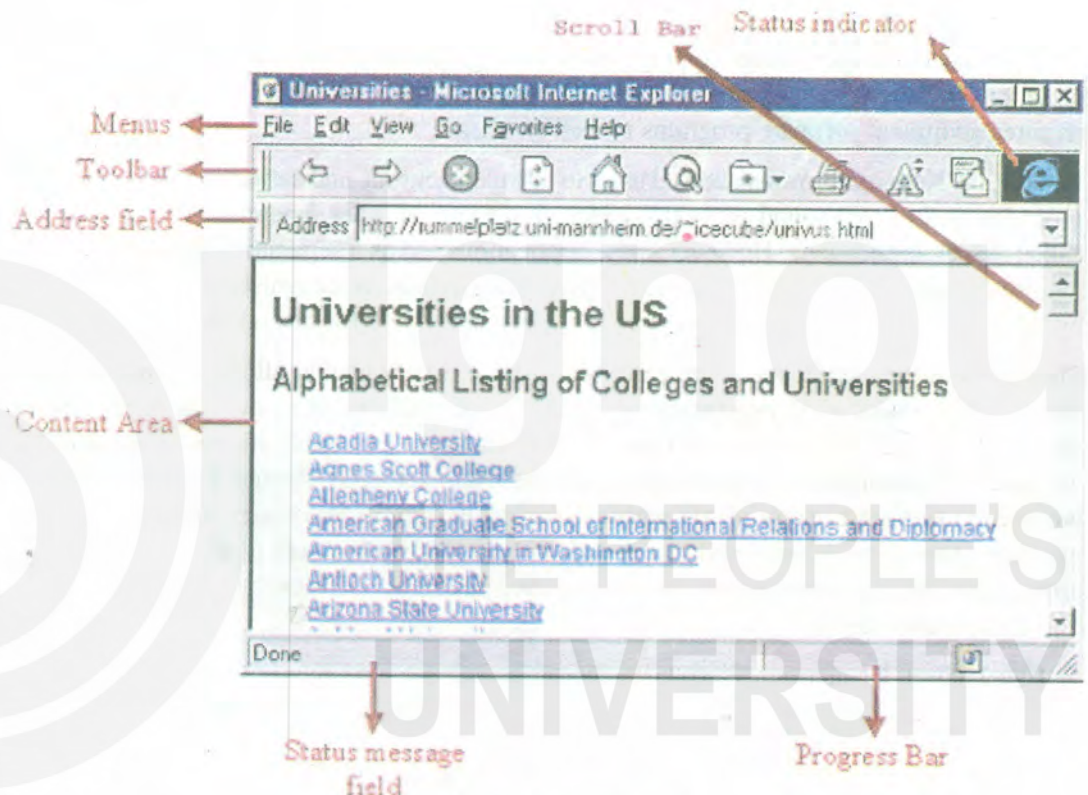


Fig. 14.1: Web Browser

Menu Bar



Menu bar in a browser exists between the title bar and the toolbar. Clicking on any of the words in the menu bar will cause a 'pop-up menu' to appear. This Unit will deal with only a couple of important menus.

File

'File' menu has options like new, open, edit, save, save as, page set-up, print, print preview, send, import and export, properties, work offline and close. Several of these options are common to other packages that students would have used. This Unit would deal with save and print menus.

Saving and Printing Web Pages

A user can save web pages on to the computer or on a disk and print out a hard copy of a page.

A user can save a web page on to computer's hard drive or on to a floppy disk. Follow the following steps in order to save a web page:

Saving a page on to the hard drive (C:)

- Open the page in the browser.
- Click on the 'File' menu
- Select 'Save As' from the file menu.
- A dialogue box will open. A user may choose the disc or folder where he / she wants to save the file, name of the file and its format.
- Click on the 'Save' button in the dialogue box.

Printing a Web Page

The first thing to ascertain when printing a web page is the size of the page. A web page may actually run into several pages. The browser would assume that a user wants to print all the contents of the document. Follow the steps given below to print only a portion of the page:

- Open the desired page in the browser.
- Click on the 'File' menu.
- Select 'Print Preview'.
- In the print preview screen, use the 'next page' button to advance till you find out total number of pages.
- Observe pages that contain the material required to be printed (The page numbers are displayed in the bottom, left corner of the screen).
- Click on the 'Close' button in the print preview dialog box.
- Either select 'Print' from the 'File' menu, or hit the 'Print' button in the toolbar.
- The print dialogue box will open.
- Enter the first page and the last page of the section of the document to be printed.
- Press 'OK'.

The document should begin to print in a moment.

File

'File' menu provides usual options such as:

- **Cut:** Cuts selected text to the clipboard.
- **Copy:** Copies selected text to the clipboard.
- **Paste:** Pastes the contents of the clipboard in other applications.
- **Select All:** Selects all text in the field where the cursor is located.
- **Find...(on this page):** Finds a text string within the document being viewed.

View

'View' menu has the following options:

- **Toolbar:** When checked, the Toolbar is displayed. If not, it is hidden from view.

- **Status Bar:** When checked, the Status Bar at the bottom of the window is displayed. If not, it is hidden from view.
- **Explorer Bar:** Explorer bar has options such as, Search, Favourite, Media, History and Folders. Most of these options are already covered under toolbar.
- **Go to:** Back, Forward and Home (same as in the toolbar).
- **Stop (Loading):** Stops or interrupts the loading of the current document. (Same as Stop Button).
- **Refresh:** Refreshes the document by reloading from local memory. (Same as Refresh Button).
- **Text Size:** Changes size of fonts to smallest, smaller, medium, larger and largest.
- **Encoding:** By default encoding is set to western European. Most web pages contain information that tells the browser what language encoding (the language and character set) to use. If the page does not include that information, and select the Auto-select feature, Internet Explorer would usually determine the appropriate language encoding.
- **Source:** Views the HTML codes for the current document.
- **Full Screen:** Hides menu bar, tool bar, address box, status bar, etc. Shows the content area on the entire screen.
- **Options:** Customises Internet Explorer for your PC.

Favourites

Favourites menu has the following options:

- **Add to Favourites:** Adds the document that is being viewed to the Favorites list.
- **Organise Favourites:** Opens the Favourites window. Allows users to rearrange their collection, delete entries, add category headers, rename items, etc.
- **Favourites Items:** Each of the Favourite items appears in this Favourites list. Select any Favourites from the menu to go to that document. Folders will allow opening of secondary menu with additional items.

Favourite is dealt in detail under the Tool Bar below.

Tools

'Tools' menu has the following options:

- **Mail and News:** Menu options allow a user to opt for the mail program of his/ her choice.
- **Synchronize:** Allows a user to update web pages from the desktop to server.
- **Windows Update:** Latest news about the Windows operating system.
- **Messenger:** Invokes Hotmail or default messenger.
- **Show Related Links:** Opens a side bar with links related to document that is being currently viewed.
- **Internet Options:** Opens the window that allows a user to customize and configure the Internet Explorer.

Help

A user can take help of the 'Help' menu as and when required. If the user has a question or a problem, or if he / she wants to learn more about the capabilities of Internet Explorer, he/she is to click on the help menu and select 'Contents and Index' which will allow the user to search through the topics of the Help menu and select appropriate help.

A browser's toolbar consists of buttons that are shortcuts for menu commands already described above. They make browsing faster and easier.



Fig. 14.2: Toolbar

The 'Back' and 'Forward' Buttons



Once a couple of web pages are loaded, 'Back' button gets activated. With a click on this button, the web browser will return the user to the last web page viewed. The 'Forward' button (when activated) takes a user to the next web page viewed.

Stop



The 'Stop' button is active (red) only when a new page is in the process of opening. While the page is opening (downloading), a user can choose to stop the download by clicking on the stop button. A user may like to stop downloading for a number of reasons, for example, the file being downloaded might be taking much more time than expected or the size of the document being downloaded is very large and is likely to take a long time to load. The 'Stop' button, like the Status Indicator, also indicates that browser is in the process of downloading information. A download is in progress when the 'Stop' button is red.

Refresh



Refresh button loads a fresh copy of the web page being displayed currently. Web pages get stored in the Internet temporary files when a user visits a website. When a user re-visits a page that he / she has visited earlier, the browser displays the file stored in cache of the PC, rather than the current page on the web to save on download time. Browser's 'Reload' button re-downloads the newest copy of the current web document.

Home



A click at the 'Home' button returns the user to the home page defined in the browser. A user can designate any web page as his / her home page.

Search



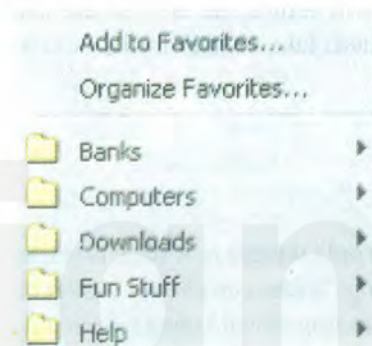
In case of Internet Explorer, a click at the 'Search' button will take a user to the MSN Search in the left pane. The Internet Explorer with its Search Assistant, gets more useful search results by specifying the type of information being searched beforehand (such as, an address, web page, company, or map). Other browsers may display a choice of popular Internet search engines in the left pane. When a link is clicked, the page appears in the right pane, so that a user does not lose the sight of his / her search results.

Favourites



Favourites (referred to as 'Bookmarks' in Netscape) provide a convenient way to keep track of commonly visited, or interesting web pages. At any time, a user can use the Favourites menu to show a list of current favourite sites and jump to a page immediately by selecting it from the menu. Favourites allow a user to save a URL of a site that he / she has visited earlier. It displays a list of the sites that a user has saved as Favourites. A click on any item in the list will trigger a visit to the selected site. A user can create, move, rename, or delete folders or files from options.

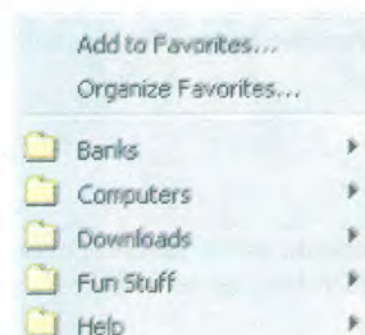
To view existing Favourites, click on the button 'Favourite' in the Internet Explorer toolbar, or bring up the Favourites menu by clicking on Favourites in the main toolbar (at the top of the screen).



To add a Favourite site in Internet Explorer, visit the web page to be added, so that it is displayed in the browser window, and its URL is displayed in the Address field. From the Favourites menu, choose the following:

To Add to Favourites....

- Edit the name of the Favourite in the Name field (if required)
- To place the new Favourite in a particular folder, click at 'Add to Favourite' button. This will display a list of available folders. Select one by clicking it once.
- Click OK.



To Create a New Favourite Folder, follow the following steps:

- Choose Organise Favourites... from the Favourite menu.
- In the Organise Favourites window, press the 'create new' button.
- The new folder will appear in the list of folders. Type a name for it (e.g. search engines, digital libraries, e-Books, etc.). Press the Enter key to confirm the name of the folder.
- Click the Close button.

To Jump to a Favourite

- Select the Favourite directly from the Favourites menu.

Organising Favourites into Folders

When a site is added to the Favourites using Add Favourite, the Internet Explorer places it at the end of the list of Favourites. As the number of Favourites grows large, specific Favourites are harder to locate. To help overcome this problem, the Internet Explorer allows users to group bookmarks together into hierarchical folders, similar to the folders used to store files on the computer's hard disk.

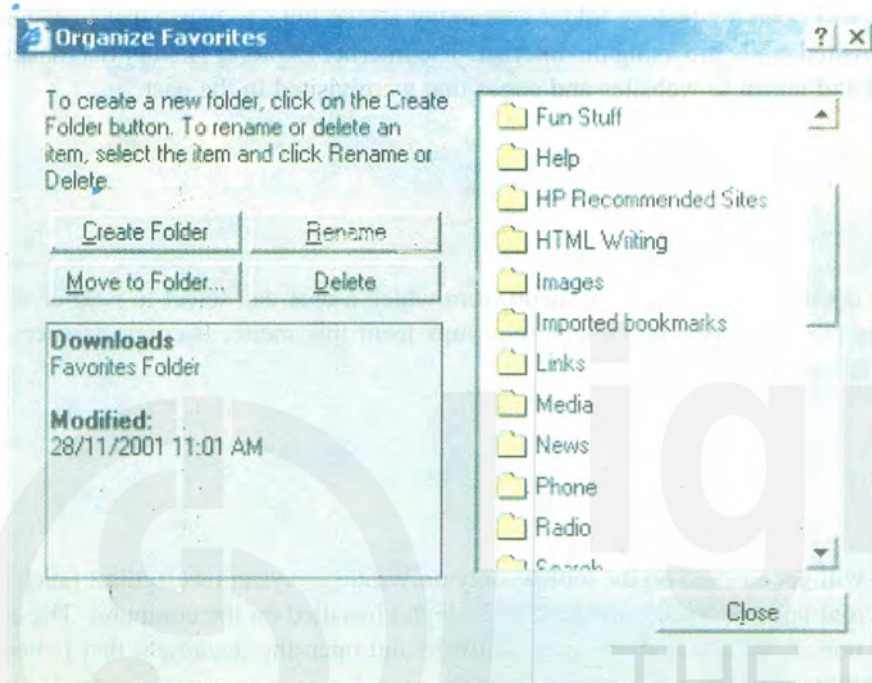


Fig. 14.3: Organising Favourite Web page

To Edit the Name or URL of a Favourite Manually

- Bring up the Favourites window by choosing Organise Favourites... from the Favourites menu.
- Select a particular Favourite by clicking on it. To edit the name of the Favourite, press the Rename button, and edit the Favourite name.
- Edit the URL associated with the favourite by clicking on the Favourite with the right mouse button, and choose Properties from the pop-up menu.
- Select the Internet Shortcut or Web Document tab and edit the URL in the URL field.
- Click the OK button. Click Close to close the Organise Favourites... window.

To Sort the Favourites in a Particular Folder

- Open the Favourites menu;
- Right-click the Favourites menu to display its context menu;
- Select Sort by Name (close to the bottom of the context menu). This will alphabetically sort this menu.

This method can be repeated on any sub-folder by right-clicking the pop-up menu of any sub-folder.

Print



Prints the page that is being viewed. This is one way to save information from the Internet and read it at leisure.

History



This button will open the history folder containing all the links to pages that have been previously visited while browsing the Internet. The Internet Explorer History list makes it easy to find and return to websites and pages that were visited in the past.

Mail



This button opens into a drop down menu from which a user can select to read or send email. Users can also open up their newsgroups from this menu. It connects users to Microsoft Outlook.

Edit



This button will get enabled on the toolbar only on Windows system web editor (such as, Microsoft Front page or Microsoft Word or Notepad installed on the computer. The edit will launch one of the above-mentioned software and open the document that is being viewed currently.

14.5.4 The Location (URL) Box



Most web browsers have an address window where a user can key-in the URL that he / she wants to visit. The URL window displays the Internet address of the website that is being viewed currently along with specific directory and filename of the current document. The location or the URL can be edited directly and a click at 'go' will take the user to the new location.

Self Check Exercise

- 5) What is a web browser? Which web browser is used the most?
- 6) What are plug-ins or helper application programs?

Note : i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

14.6 MARK-UP LANGUAGES

The mark-up is used in printing industry as instructions for printing in a particular style. Mark-up is also used while proof reading, editors mark the text as instructions to the printer to print the text in a particular font type, size and in bold or italics. Similarly, to display the electronic text in a web page instructions are given as mark-up within the text to make parser (a computer program) understand how the text should appear on display. Mark-up is also used for data retrieval, particularly in the library and information field. Once the structure of document is fixed, one can easily find which part of the document contains which kind of data.

Three basic concepts are fundamental to understanding of all mark-up languages, when described in SGML terms. These are: i) a mark-up entity, ii) a mark-up element and its associated attributes; and iii) a document type.

Entity in SGML is text that is composed of streams of symbols (characters or bytes of data, marks on a page, graphics, etc.). At a higher level of abstraction, a text is composed of representations of objects of various kinds, linguistically or functionally defined. Such objects do not appear randomly within a text, but various types of objects appear in specifiable relationship to other objects, i.e., they may be included within each other, linked to each other by reference or simply presented sequentially. This level of description sees text as composed of structurally defined objects, known as elements in SGML. The grammar defining how elements may locally be combined in a particular class of texts is known as a document type. These three fundamental concepts are adequate to describe all the complexities of marked-up texts, of whatever kind and for whatever purposes.

14.6.1 Standard Generalised Mark-up Language (SGML)

SGML is application independent, non-proprietary and extremely flexible mark-up language. It was first developed in 1970 as GML (Generalised Mark-up Language) and evolved into an International Standard (ISO, 1986). SGML is frequently referred to as a meta-language, which means that SGML is not a single language but a language that describes a family of mark-up languages. In other words, SGML is the framework for defining particular mark-up languages. SGML is an effective solution for handling complexity of electronic publishing because of its powerful and flexible structuring capabilities, as well as for its capacity to capture and organise information about the publications (metadata). It provides for descriptive, as opposed to procedural mark-up. That is, it simply, states names to categorise parts of a document instead of specifying process to be carried out.

SGML uses text characters both for the text as well as for mark-up that describes that text. It has no proprietary codes; instead, each user (or group of users) may create whatever codes are necessary and meaningful for what is being published. A publisher can define his own set of codes for books and journal publishing. The key to self-defined codes in an SGML document is called DTD (Document Type Definition). Code sets or DTDs can be specific to a single book or journal or can span to a group of related books or journals. An SGML document consists of the following three distinct parts:

Declaration: It gives fundamental information like language of document and code set being used (i.e., English/ASCII).

DTD: Details of codes and rules restricting their use.

Instance: The text being published, marked up with the codes described in the DTD.

SGML concerns itself with the structural features of a document while the appearance and display features are left to the ultimate presentation system to determine how those features appear on display or print. Resultantly, when documents move from system to system, or portions of one document are used in another, they do not need to be recoded. Because of its powerful and flexible structuring capabilities, as well as its capability to capture and organise information about the publications, SGML-coded documents can be

used effectively to search information contents of documents based on the structure and content of the information. Many SGML depositories are considered as 'text databases'. Since they enable a publisher to organise the published information in different ways for different contexts.

Contents of an SGML document are stored separately from its format, resultantly contents or parts of contents can be rendered in different ways for different needs, platforms and display methods. SGML is often used as an archival format and for document reuse and repurposing. Richly coded SGML documents also facilitate more complex searching than unstructured, word-processed text. For fully marked-up documents, searches can be made on bibliographic citation marked <site> or such citations can be extracted from each document to create a citation database as a secondary product.

SGML liberates documents from the cumbersome and costly process of conversion from system to system. It does not require any special hardware or software. It is possible to create a valid SGML file in any word processor or text editor although there are a number of SGML-based systems available in the market. SGML preserves the document and its coding from obsolescence as well. Owing to the fact that an SGML document incorporates the key to its own codes (Declaration, DTD), it is possible to validate SGML codes by parsing SGML file. Parsing is a process by which the document instance is checked against the declaration and the DTD to make sure all the codes in a file are legal and used properly.

14.6.2 Extensible Mark-up Language (XML)

XML is a subset of the Standard Generalised Mark-up Language (SGML). It is designed to make it easy to interchange structured documents on the web. The Cascading Style Sheet (CSS) developed for HTML can also function for XML to take care of formatting and appearance. Unlike HTML, XML allows for the invention of new codes. XML files are not only consistent and compatible with SGML; it also simplifies SGML in many ways. For example, while SGML allows 'tag minimisation', enabling the omission of end tags, XML always requires explicit end tags that make it a lot easier to write tools and browsers. XML introduces the concept of a 'well-formed' document, one in which the tags used are nested correctly and proper XML syntax is followed. In addition, like SGML, XML allows for 'valid' documents too, which go a step beyond 'well formed' status by using an explicit structure defined in a DTD. 'Well-formedness' is a very appealing feature of XML, because it allows publishers to tag what they are publishing in whatever way is meaningful, without being confined to a specific set of tags (as with HTML) or needing to write a DTD.

An XML document may require companion XSL (Extensible Style Language) to reformat it into RTF, LaTeX or any other format. The XSL also makes it possible to offer database functionality from XML documents with no actual database needed. XML also defines how Internet Uniform Resource Locators can be used to identify component parts of XML data streams. Akin to an SGML document, XML documents can also be verified to ensure that each component of document occurs in a valid place within the interchanged data stream by defining the role of each element of text in a formal model, known as a Document Type Definition (DTD). An XML DTD allows computers to check, for example, that users do not accidentally enter a third-level heading without first having entered a second-level heading, something that cannot be checked using the Hypertext Mark-up Language (HTML). However, unlike in SGML, DTD is not a necessity in XML. If no DTD is available, either because all or part of it is not accessible over the Internet or because the user failed to create it, an XML system can assign a default definition for undeclared components of the mark-up.

XML allows users to:

- bring multiple files together to form compound documents;
- identify where illustrations are to be incorporated into text files, and the format used to encode each illustration;

- provide processing control information to support programs, such as, document validators and browsers;
- add editorial comments to a file.

Like SGML, XML does not have a predefined set of tags, of the type defined for HTML that can be used to mark-up documents in a standardised template for producing particular types of documents. XML is a formal language that can be used to pass information about the component parts of a document to another computer system. XML is flexible enough to be able to describe any logical text structure, whether it be a form, memo, letter, report, book, encyclopaedia, dictionary or database.

XML is based on the concept of documents composed of a series of entities or objects. Each entity or object can contain one or more logical elements. Each of these elements can have certain attributes (properties) that describe the way in which it is to be processed. XML provides a formal syntax for describing the relationships between the entities, elements and attributes that make up an XML document, which can be used to tell the computer how it can recognise the component parts of each document. XML differs from other mark-up languages in that it does not simply indicate where a change of appearance occurs, or where a new element starts. XML sets out to clearly identify the boundaries of every part of a document, whether it is a new unit, or a reference to another publication. The structure of a document can be checked if the user provides a document type definition that declares each of the permitted entities, elements and attributes, and the relationships between them.

14.6.3 Hypertext Mark-up Language (HTML)

Hypertext Mark-up Language (HTML) is an SGML application complete with DTD. It is designed to tell a browser how to display documents on the web. HTML is the *de facto* language of the web and is largely responsible for resurgence of interest in SGML in the past few years. Unlike SGML, HTML has a pre-defined set of codes, that are easy to learn and use and build tools for writing HTML pages. HTML codes are embedded into the text that communicate to a web browser such as, Netscape Navigator or Microsoft Internet Explorer. Like SGML, it also uses simple text or ASCII for text as well as for the HTML codes. An HTML page can thus be built using a word processing package or a text editor. There are several HTML editors and conversion programs that act similar to a word processing package. These editors typically show the codes as they are inserted. In a What You See Is What You Get (WYSIWYG) environment, such as, MS Word or other MS Windows packages, the user never sees these codes. Web browsers are similar to WYSIWYG word processors because they read the embedded codes and then apply them to specified texts.

HTML is competent at presenting text, graphics, images in a reasonably decent layout. web browsers readily accommodate a multitude of plug-ins that allow inclusion of audio, video, 3-D and other specialised files. Any of these can also be included as a link in a standard HTML page. Clicking the link loads the plug-in to view or play the file. HTML files are tiny since they are simple text files. Further, the static HTML web pages can be transformed into vibrant, dynamic and interactive web creations using ever-evolving web technologies like CGI Script, Perl, Java, Javascript, ASP, DHTML, XML and Open Database Connectivity (ODBC) for incorporating interactivity on a web page.

Simplicity of HTML is also its serious limitation for books and journals. HTML does not provide enough codes to present complexities of a scientific text. HTML does not provide for Greek and mathematical characters that are important to scientific text. Moreover, HTML is all about presentation and not for structure or contents. The only contents that it successfully describes is in metadata codes or in its title. Furthermore, an HTML file can be derived from an SGML file any time but the reverse is not possible. The competency of HTML at presenting text has further been enhanced with use of Cascaded Style Sheet (CSS).

Editor for HTML

HTML is a plain text file and needs a simple text editor to create the tags. However, it is important that all HTML documents have the extension .html or .htm which is three / four letter extension. Windows 'Notepad' may be used as an editor for writing the HTML files.

Syntax of HTML Commands

HTML documents are plain text documents (ASCII) and contain: i) the actual text of the document, and ii) the HTML tags. In general, HTML tags look like this:

<TAG>The text that will appear in the browser</TAG>

HTML tags consist of a left angle bracket (<), a tag name, and a right angle bracket (>). Tags are usually paired (e.g., <HTML> and </HTML>) to start and end the instruction. The end tag looks similar except for the slash (/) that precedes the tag name. HTML is not case sensitive, i.e., <html> is equivalent to <HTML> is equivalent to <hTmL>. It is a good practice to put HTML tags in caps to make it easier to read and edit documents because the capital letters would stand out from the text.

All HTML tags do not have a beginning and an end tag. Some tags are one-sided. A common one is the line break element which breaks a line of text at the point where it appears -
.

Structural Elements of a Document

Every HTML document should contain certain standard HTML tags. These tags describe the overall structure of a document, identify the document to browsers and provide simple information about the document. These structure tags do not affect its appearance and they are essential for tools that interpret HTML files. These structural elements are:

The html tag, <HTML> ... </HTML>

The head tag, <HEAD> ... </HEAD>

The body tag, <BODY> ... </BODY>

- <HTML> ... </HTML>

The <HTML> tag indicates that the content of the file is in the HTML language. All the text and commands in a document should go within the beginning and ending HTML tags.

```
<HTML>
... the document ...
</HTML>
```

- <HEAD> ... </HEAD>

The <HEAD> tag specifies a limited amount of bibliographic data related to the document. It is the second item in the document. This element does not contain any text that displays in the browser except for the title that appears in the title bar of the browser.

```
<HTML>
<HEAD>
<TITLE>
... title information ...
</TITLE>
</HEAD>
... the document ...
</HTML>
```


- `<TITLE> ... </TITLE>`

Each HTML document needs a title to describe the content of the document. The title is used by the browser to display it in its title bar and is commonly indexed by search engines such as, Google, Yahoo, WebCrawler, etc. To give a document a title, use the `<TITLE>` tag. This tag always appears inside the `<HEAD>` element.

```
<HTML>

<HEAD>

<TITLE>A unique title</TITLE>

</HEAD>

... the document ...

</HTML>
```

Do not use any other tags within the `<TITLE>` tag.

- `<BODY> ... </BODY>`

The `<BODY>` tag follows the `HEAD` tag. It contains all parts of the document to be displayed in the browser

```
<HTML>

<HEAD>

<TITLE>A unique title</TITLE>

</HEAD>

<BODY>

... the document ...

</BODY>

</HTML>
```

- A Note About Blank Space

When an HTML document is 'read' by a browser, any extra spaces, tabs, returns, etc. that are added to the document are ignored. The only thing that formats an HTML document is an HTML tag. This feature of mark-up language is used to provide more prominence to the editable text keeping the tags at a distance.

- Heading

Headings are used to divide sections of text, like in any document. They are used to designate the logical hierarchy of the HTML document. There are currently six levels of headings defined. The number indicates heading levels (H1 through H6). Each heading, when displayed in a browser, will display differently. Headings should be used in descending order to indicate logical structure. If only two headings are being used, they should be `<H1>` and `<H2>`

Indention in the example given below is essentially to show the hierarchy. As mentioned above, blank space in an HTML document does not have any significance.

```
<H1>Animal Kingdom</H1>

    <H2>Vertebrates</H2>

        <H3>Mammals</H3>
```


<H3>Reptiles</H3>

<H2>Non-vertebrates</H2>

<H3> Arthropoda </H3>

<H4>Insects</H4>

<H4>Spiders</H4>

<H4> Crustaceans </H4>

<H3>Mollusca</H3>

The title of the document is often repeated as a Heading I. While the <TITLE> of the document appears in the title bar of the browser, the <H1> heading appears in the text. Do not use headings to format text. Heading tags are used to provide structure to a document.

● Paragraphs

Unlike documents in word processors, carriage returns in HTML files are not significant. Use paragraph tag <P> to indicate a paragraph. A browser ignores any indentations or blank lines in the source text. Without a <P> tag, the document becomes one large paragraph. The paragraph tag indicates a plain text paragraph. The closing tag </P> is optional. However, many browsers expect the opening paragraph tag to indicate the end of a paragraph.

<P>

Text for the paragraph.

</P>

If paragraph alignment attributes are included, the end tag must be used.

<P ALIGN=RIGHT>

Text for the paragraph aligned to the right

</P>

● Lists

Lists are one of the most common elements of HTML that are used very often. The HTML supports the following five types of lists:

Unordered, or bulleted lists: Labelled with bullets or some other symbol (○, ■, ●)

Ordered, or numbered lists: Labelled with numbers, i.e., 1 .., I ... ; i A .., a ...)

Definition, or glossary lists: Each item has a term and its definition arranged like a definition list or like a glossary list.

All list tags have the following common elements:

The entire list is surrounded by appropriate opening and closing tags to identify the type of list. For example:

 and for an unordered list

 and for an ordered list

<DL> and </DL> for glossary list

Each item within the list has its own tag: <DT> and <DD> for the glossary lists and for ordered and unordered lists.

• Unordered and Ordered Lists

Unordered Lists

Unordered lists or bulleted lists are lists in which the elements can appear in any order. The browser inserts a bullet or some other symbol in front of each line. Example of an Unordered List is given below:

```
<UL>
<LI>Classification</LI>
<LI>Cataloguing</LI>
<LI>Serials Control</LI>
</UL>
```

Ordered Lists

Ordered lists are used to indicate specific order of elements in a web document. When the browser recognises an ordered list, it numbers (and often indents) each of the elements sequentially. The browser automatically does numbering. An example of an Ordered List is as follows:

```
<OL>
<LI>Morphology</LI>
<LI>Anatomy </LI>
<LI>Physiology </LI>
</OL>
```

Definition Lists

Definition lists are slightly different from other lists. Each list item in a definition list has two parts:

- A definition term <DT>
- Term's definition <DD>

Both <DT> and <DD> are one-sided tags. The entire definition list is indicated by the tags <DL>... </DL>. Example of a Definition List is given below:

```
<DL>
<DT>Anatomy
<DD> The science of understanding the structure and make-up of
the body
<DT>Morphology
<DD> Morphology is the science of the structure of animals and
plants <DT>Physiology
<DD> The study of the physical and chemical processes involved in
the functioning of the human body
</DL>
```


Character Formatting

When HTML tags are used for paragraphs, headings and lists, those tags affect the text as a whole. Character styles are tags that affect words or characters within other HTML entities and change the appearance of that text; for example, making it bold or italics.

To change the appearance of a set of characters, two kinds of tags can be used: logical style tags or physical style tags. It is advisable to use logical styles instead of physical styles to mark up the documents. Future releases of HTML might not support physical styles, which could mean that browsers will not display physical style coding.

Logical Style Tags

Logical style tags indicate how the text is to be used, not how it is to be displayed. It is for the browser to determine the actual way the text within these tags is presented. For example, use of logical tags will ensure that text sandwiched between logical text in given prominence. However, difference browsers may determine how prominence has to be given. There are presently eight logical style tags:

`<EM> ... </EM>`: for emphasis. Typically displayed in italics.

`<STRONG> ... </STRONG>`: for strong emphasis. Typically displayed in bold.

`<CODE> ... </CODE>`: for computer code. Displayed in a fixed font.

`<CITE> ... </CITE>`: for titles of books, films, etc. Typically displayed in italics.

- `<KBD> ... </KBD>`: for user keyboard entry. Typically displayed in plain font.

- `<SAMP> ... </SAMP>`: for a sequence of literal characters. Displayed in a fixed font

- `<VAR> ... </VAR>`: for a variable. Typically displayed in italics.

- `<DFN> ... </DFN>`: for a definition. Typically displayed in italics.

- Seldom used style tags. They are supported by HTML Level 2 specifications, but are not handled uniformly by browsers.

Physical Style Tags

Physical style tags indicate exactly the way the text is to be formatted. Physical style tags are, however, browser dependent. If a certain browser cannot display one of the physical styles, it may substitute another style. There are four physical style tags:

`<B> ... </B>`: for bold

`<I> ... </I>`: for italic

`<TT> ... </TT>`: for typewriter font - teletype

- `<U> ... </U>`: for underline

- Use of the underline element is discouraged because of poor support by browsers and the fact that underlining is often used to indicate hypertext links.

- **More HTML TAGS**

Horizontal Rules - `<HR>`

The horizontal rule tag, `<HR>`, has no closing tag and no text associated with it. The `<HR>` tag creates a horizontal line on the page. It is excellent for visually separating sections on your web page. It is often seen at the end of text on web pages and before the address information.

Example:

Document text. .

<HR>

More document text > . .

Line Break -

The line break tag,
, breaks a line of text at the point where it appears. Just like the <HR> tag, it does not have a closing tag.
 does not add extra space above or below the line. All it does is restart the text at the next line.

Example:

Indian Institute of Technology Delhi

Haus Khas

New Delhi 110016

INDIA

Addresses - <ADDRESS>

The address tag, <ADDRESS>, usually goes at the bottom of the page and indicates who wrote the web page, whom to contact for more information, the date, any copyright notices or anything else that seems appropriate. The <ADDRESS> is often preceded by a horizontal rule and the line break tag can be used to separate the lines.

Example:

<ADDRESS>

Indian Institute of Technology Delhi

Hauz Khas

New Delhi 110016

Preformatted Text - <PRE>

The preformatted text tag, <PRE>, is most commonly used to create tabular data and preserving the white space that you use, for example, spaces and tabs. It is also excellent for things like code.

Example:

<PRE>

text.

text

text

text

text

</PRE>

● Linking

The power of HTML comes from its ability to link text and / or images to another document or section of a document. Browsers highlight identified text or images with colour and / or underlining to indicate that it is a hypertext link.

To create a link, the following two things are required:

- the name and location (path) of the file to be linked
- the text that will be highlighted in the browser
- **Absolute Pathnames Versus Relative Pathnames**

Pathnames identify the location of the files. An absolute pathname is the entire path of the file starting at the root. It is the actual location of the file, for example, `c://documents/library/services.html` is an example of absolute path name.

A relative pathname describes the file in relation to the current directory. Relative pathnames can include directories and might also include directions, for example, to go up two directory levels, and then go down two other directories, etc.

One should use relative pathnames in preference to absolute pathnames. Absolute pathnames may seem easier, but are not portable. Specifying relative pathnames allows movement of documents from one directory to another on from one system or to another system, with very little or no modifications needed. It is much easier to maintain HTML documents with relative pathnames.

- **Uniform Resource Locator (URL)**

The World Wide Web uses Uniform Resource Locators (URLs) to specify the location of files. A URL includes:

- the protocol being used (e.g., ftp, Gopher, etc.), for example: **http://**
- the host name, for example: **www.iitd.ac.in**
- the port number (generally omitted, unless otherwise specified)
- and the directories and file name, for example: **/acad/library/index.html**

The URL would look like this:

`http://www.iitd.ac.in/acad/library/index.html`

HTTP stands for HyperText Transfer Protocol and is the protocol that the World Wide Web servers use to send HTML documents over the Internet.

- **The Link Tag - <A>**

The link tag, <A>, (also called anchor tag) is used to create links in an HTML document. It is also used to create anchors for links.

The <A> tag has some extra features. The opening tag includes both the name of the tag, (A) and attributes of the tag. The most common anchor attribute is a Hypertext REFerence (HREF). The HREF attribute is used to specify the name (pathname) or URL of the file where this link points. The link tag also has a closing tag, . All of the text between the opening and closing tags becomes the actual link on the screen and is highlighted. That's the text that a user clicks so as to jump to the document specified by the HREF attribute.

- **Absolute Link**

URL of Remote File

Link that appears in browser

`<A HREF=http://www.iitd.ac.in/">IIT Delhi Home Page </A>`

- **Relative Pathname**

Relative Pathname **Link that appears in browser**

`<A HREF="files/home/ehome.html">Home Page</A>`

Opening Tag

Closing Tag

Another common attribute is the NAME attribute. The NAME attribute takes a keyword that will be used to reference the anchor. It is useful to create a link within a document.

Anchor name to **Text that will be at the top of the screen in the browser**

`<A NAME="Section5">The Fifth Section</A>`

Opening Tag

Closing Tag

- **To link to this anchor**

Anchor name to link **Link that appears in browser**

`<A HREF="#Section5">The Fifth Section</A>`

Opening Tag

Closing Tag

The # character indicates that the target of the link is within a document.

- **Embedding Images**

The World Wide Web is a multimedia information system. The HTML format permits images to be embedded into your document. The tag does not have a closing tag.

- **Inline Images**

Inline images are images that appear next to the text in an HTML document. Images may be in .gif or .jpeg formats. Don't overload documents with images. Each image takes time to load and will slow down the length of time it takes for a document to display. It is recommended that the text and images used on a page should be less than 50k and never more than 100k.

Example:

`<IMG SRC=image.gif BORDER=1 WIDTH=19 HEIGHT=19 ALT=" ">`

- **Image Attributes**

Image attributes refers to the different ways you can present the image, i.e., its size, location of the text and providing alternate text.

Table 14.1: Attentative to Present the Image

Image Attributes	DESCRIPTIONS
SRC	Indicates the source location (i.e., path and filename) of the image.
ALT	Designates alternate text that will be inserted in place of the image (for text-based browsers or images that do not load properly).
ALIGN	Aligns the specified text with the related image-either top, bottom or centre.
WIDTH	Indicates the width of the image in pixels.
HEIGHT	Indicates the height of the image in pixels.
HSPACE	Indicates the space on the left or the right of the image in pixels.
VSPACE	Indicates the space above and below the image in pixels.

ALT Attribute

Some World Wide Web browsers cannot display images, they display text only. Some users turn off image loading even if their software can display images (especially if they are using a modem or have a slow connection). HTML provides a mechanism to give readers a little information about the images they cannot view. The ALT attribute lets the reader specify text to be displayed instead of an image.

To display alternate text for an image:

```
<IMG SRC=image.gif ALT="Beach Scene"/>
```

It is recommended to include alternate text for each image in a document.

ALIGN Attribute

There is some flexibility in displaying images. You can have images separated from text and aligned to the left or right. Or you can have an image aligned with text. Try several possibilities to see how your information looks best.

By default, the bottom of an image is aligned with the text. You can also align images to the top or centre of a paragraph using the ALIGN= attributes TOP and CENTER

To centre your Image:

```
<IMG SRC=image.gif ALIGN=CENTER>
```

Size attributes - HEIGHT and WIDTH

Image size attributes tell your browser the size of the images it is downloading with the text. The HEIGHT and WIDTH attributes let your browser set aside appropriate space (in pixels) for the images as it downloads the rest of the file. It is highly recommended that you always specify height and width.

```
<IMG SRC==image.gif HEIGHT=100 WIDTH=65>
```

Space attributes - HSPACE and VSPACE

You can also leave extra space between the text and the image by using the HSPACE== and the VSPACE= attributes.


```
<IMG SRC==image.gif HSPACE=15>
```

• External Images

You may want to have an image open as a separate document when a user activates a link on either a word or a smaller, inline version of the image included in your document. This is called an external image, and it is useful if you do not wish to slow down the loading of the main document with large inline images.

The same steps are taken to link to an image as to a file.

Example:

```
<A HREF="image.gif">Text in browser</A>
```

You can also use a smaller image as a link to a larger image. Usually, the smaller image will be the .gif file and the larger image would be a.jpg file.

```
<A HREF="large.jpg"><IMG SRC="small.gif"></A>
```

The reader sees the small.gif image and clicks on it to open the large.gif file.

• Special Characters

There are some characters that cannot be used as 'literals' in HTML. The 'less than' (<) and 'greater than' (>) symbols are good examples because HTML reads them as delimiters for tags. These 'special characters' need to be treated differently when they are to appear as text on a web page.

Common HTML special characters include:

< > & " (C) (R)

To insert these characters in HTML documents, a string of literals known as entities must be used. Entities begin with an ampersand (&) and end with a semi-colon (;). For the ISO-Latin-1 character set (ISO8859-1), a numeric entity set has been devised. For common symbols and letters, an equivalent character entity can also be used. Here are some examples of how to enter special characters in HTML:

	Character		Numeric Entity		Character Entity (if available)
For	<	Use	<	Or	<
	&		>		>
	&		&		&
	(c)		©		©
	(R)		®		®
	A		À		À
	e		é		é
	1/4		¼		(N/A)
	<<		«		(N/A)
	>>		»		(N/A)

● Tables

Tables are useful when information is to be displayed in a certain way. Tables have several tag definitions, which can be modified, in any number of ways to get the desired effect. A table begins with the <TABLE> tag and ends with the </TABLE> tag.

Within the <TABLE> tag designate:

- What size of border is required around the table (if any) (BORDER=#)
- The number entered here would indicate the thickness of table border in terms of number of pixels.
- How much space is required within each individual cell of table (CELLPADDING=#)
- How crowded each cell can be?
- The colour of the background of the table, can also be changed, if desired.
- Requirement of visible dividers between each cell (CELLSPACING=#)
- This number indicates the thickness of cell divider in terms of number of pixels. Usually, it looks best if it matches the number you've chosen for your BORDER.
- The rows across the table are designated by the <TR> and the </TR> tags. Between these tags, individual cells are designated with the <TD> and the </TD> tags.

With just these few tags, a very basic table can be constructed. For example, the HTML for a table of 2 rows across with 3 columns down would look like this:

```
<TABLE BORDER=3 CELLPADDING=3 CELLSPACING=3>
<TR> <TD>Cell 1</TD>
<TD>Cell 2</TD>
<TD>Cell 3</TD> </TR>
<TR> <TD>Cell 4</TD>
<TD>Cell 5</TD>
<TD>Cell 6</TD> </TR>
</TABLE>
```

The above-mentioned HTML Tags would produce the table given below:

Cell 1	Cell 2	Cell 3
Cell 4	Cell 5	Cell 6

If a cell is required to span across more than one column (i.e., join cells), add in the <TD> tag COLSPAN=#. The number entered would indicate the number of columns to span. Similarly, a cell can span to more than one row by specifying command ROWSPAN=# to the <TD> tag.

For instance:

```
<TABLE BORDER=3 CELLPADDING=3 CELLSPACING=3>
<TR> <TD>Cell 1</TD> <TD COLSPAN=2>Cells 2 and 3</TD> </TR>
<TR> <TD ROWSPAN=2>Cell 4</TD> <TD>Cell 5</TD> <TD>Cell 6</TD> </TR>
<TR> <TD>Cell 7</TD> <TD>Cell 8</TD> </TR> </TABLE>
```


The above-mentioned HTML Tags would produce the table given below:

Cell 1	Cell 2 and 3	
Cell 4	Cell 5	Cell 6
	Cell 7	Cell 8

Background and font colour can also be changed within the cells.

• Frames

Frames are a technique used in web pages to divide the page into multiple windows, where each window is called a frame and can contain its own separate page. The advantage of frames is that one window can be scrolled or changed while other windows remain fixed for such purposes as keeping a menu in view all the time. The disadvantage is that not all browsers support them. Tags used for making a frame are as follows:

`<frameset></frameset>`

Precedes the `<body>` tag in a frames document; can also be nested in other framesets

`<frameset rows="value,value">`

Defines the rows within a frameset, using number in pixels or percentage of width

`<frameset cols="value,value">`

Defines the columns within a frameset, using number in pixels or percentage of width

`<frame>`

Defines a single frame within a frameset

`<noframes></noframes>`

Defines what will appear on browsers that don't support frames

14.6.4 Dynamic HTML

The Dynamic HTML is a collective term for a combination of the new Hypertext Mark-up Language (HTML) tags and options, that will let you create web pages more animated and more responsive to user interaction than the 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 colour 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' web browser, Navigator 4.0 (and upwards) (part of Netscape's Communicator suite), and by Microsoft's browser, Internet Explorer 4.0 (and upwards). 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 website must create two versions of each site and serve the pages appropriate to each user's browser version.

The Concepts, and Features in Dynamic HTML

Both Netscape and Microsoft support:

- An object-oriented view of a web page and its elements
- Cascading style sheets and the layering of content
- Programming that can address all or most page elements
- Dynamic fonts

Style Sheets and Layering

A style sheet describes the default style characteristics (including the page layout and font type style and size for text elements such as, headings and body text) of a document or a portion of a document. For web pages, a style sheet also describes the default background colour or image, hypertext link colours, and possibly the content of the page. Style sheets help to ensure consistency across all or a group of pages in a document or a website.

Dynamic HTML includes the capability to specify style sheets in a 'cascading style sheet' fashion (that is, linking to or specifying different style sheets or style statements with predefined levels of precedence within the same or a set of related pages). As a result of user interaction, a new style sheet can be made applicable and result in a change of appearance of the web page. You can have multiple layers of style sheet within a page, a style sheet within a style sheet within a style sheet. A new style sheet may only vary one element from the style sheet above it.

Layering is the use of alternate style sheets or other approaches to vary the content of a page by providing content layers that can overlay (and replace or superimpose on) existing content sections. Layers can be programmed to appear as part of a timed presentation or as the result of user interaction. In Internet Explorer 4.0, Microsoft implements layers through style sheets. Netscape supports the style sheet approach but also offers a new HTML tag set (that Microsoft does not support). The W3C Working Committee is considering both approaches and both companies say they will support whatever W3C decides and that will be the recommended approach.

14.6.5 Virtual Reality Modelling Language (VRML)

VRML, often pronounced 'ver-mull,' is Virtual Reality Modelling 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 that you can download for the Windows platforms are blaxxun's CC Pro, Platinum's Cosmo Player, webFX, WorldView, and Fountain. Whirlwind 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, inline images let you include graphics from anywhere on the web on to your web page. In VRML, you can have inline parts of a virtual world, so that a chair in a VRML world can come from a URL on a server in

France, while the garden comes from a server in Japan, and the soundtrack is from 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 visualisation
- Distributed, multi-user environments
- Social computing
- User interfaces to information
- Financial applications
- Product marketing and advertising
- Education
- Entertainment

Self Check Exercise

- 7) Define mark-up languages.
- 8) What are the major differences between HTML and XML?

Note : i) Write your answers in the space given below.
 ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

14.7 INTERNET APPLICATIONS

Internet is the largest, vast and most complex and unorganised learning source in the world. The Internet hosts knowledge resources virtually in all disciplines. Besides, the Internet facilitates communication quickly and effectively with others who are also interested in the same discipline. Teachers, researchers, students, and other educators can share ideas instantly across vast distances.

Individuals, companies, and institutions use the Internet in 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 audio and

video conferencing and other forms of communication that allow people to telecommute or work from a distance.

The use of Internet services has resulted in increase in communication among companies, co-workers, and individuals. Media and entertainment companies use the Internet to broadcast audio and video, including live radio and television programmes; to offer online chat, online news and weather information. 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 services.

A variety of Internet tools/ services have been developed over the years to make effective use of the resources and communication capabilities offered by the Internet. These can broadly be grouped into the following four categories:

- a) Internet-based Communication Services
 - i) Communication among individuals: electronic mail, Internet telephony, and Internet chat.
 - ii) Communication Services for groups: Internet relay chat, electronic discussion groups, Listserv, News groups, Usenet and video conferencing.
- b) Connectivity
 - i) Telnet
 - ii) Remote Login
 - iii) File Transfer (FTP)
- c) Information Resources
 - a) Gopher
 - b) WWW
 - c) FTP Servers
- d) Searching information resources on Internet: WAIS, search engines and web directories.

While the World Wide Web (WWW) has already been described in this Unit, search engines and web directories are dealt in detail in Unit 13 of this course. Other Internet services are discussed in brief in this Unit.

14.7.1 Internet-based Communication Services

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. E-mail is the term given to an electronic message, usually in the form of simple text message, that a user types at a computer system and transmits it over some form of computer network to another user, who can read it. Email offers speedy and economical transfer of messages anywhere in the world. Sending e-mail messages is virtually free even to long-distance destinations. The e-mail offers the following benefits over traditional paper-based memos and postal systems:

- Messages can be sent at any time across the world as easily as across the office, to a group of people or a single recipient, without the sender leaving their desk. Messages can be logged, ensuring some form of record is held, and messages are stored when the recipient is away from his/her desk.

- The recipient can collect the mail whenever he/she wants, from wherever they are. Mobile users can collect their mail while on the move, or at other locations.
- The addressee gets the mail directly, without passing through any third party.
- Unless printed, email messages require no paper or resources other than storage space on a computer disk drive.

E-mail Client

An e-mail client is an application that is used to read, write and send e-mail. In simple terms it is the user interface to the e-mail system. The client usually consists of a combination of a simple text editor, address book, filing cabinet and communications module. The text editor allows for the creation of the message itself, and usually includes simple spell checking and formatting facilities.

The e-mail clients also facilitate files, documents or pictures to be attached to the message. The address book allows the users to store commonly used e-mail addresses in an easy to get format, reducing the chance of addressing errors. The filing cabinet allows for the storage of e-mail messages, both sent and received, and usually provide some form of search function, allowing easy retrieval of a desired message.

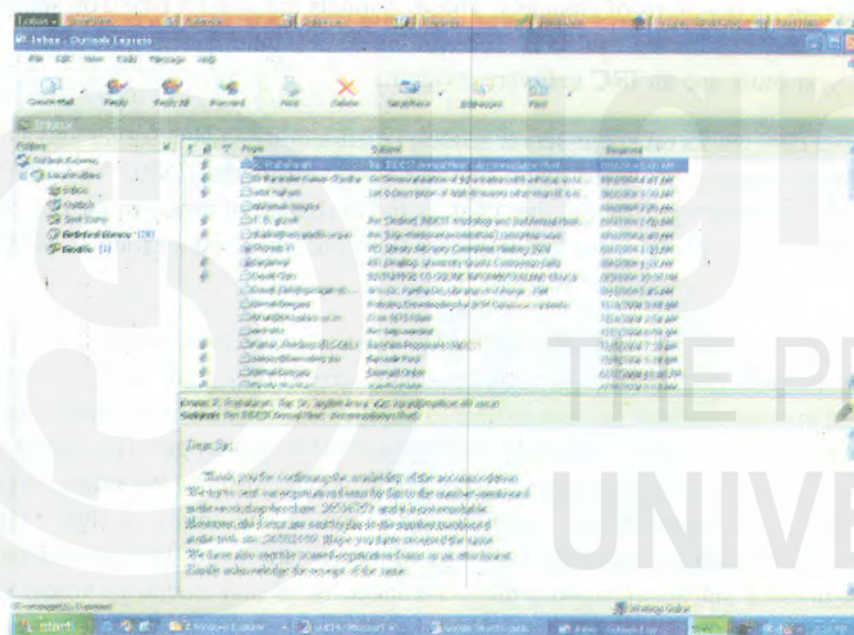
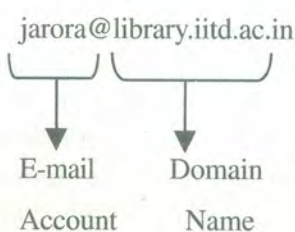


Fig. 14.4: Microsoft Outlook Express: E-mail Client

What is a Mail Server?

A mail server is an application that receives e-mail from e-mail clients or other mail servers. It is the workhorse of the e-mail system. A mail server usually consists of a storage area, a set of user definable rules, a list of users and a series of communication modules. The storage area is where mail is stored for local users, and where messages that are in transit to another destination are temporarily stored. It usually takes the form of a simple database of information. A person, sometimes called a Postmaster, maintains the mail server and the list of user accounts that it supports.

An example of a typical e-mail address looks like this:



This address is made up of two parts:

E-mail Account	The email account 'jarora' exists on mail server at iitd.ac.in under group 'library'
Domain Name	iitd is the name of institute with domain 'ac'

Chat or Instant Messengers

Chat refers to any live discussion conducted using the Internet, usually between more than two persons using their keyboard to communicate. Chat programs allow users on the Internet to communicate with each other by typing in real time. It is a feature offered by many online services or websites that allow users to 'chat' by typing messages which are displayed almost instantly on the screens of other users who are using the chat room at a given time. After entering a chat room, any user can type a message that will appear on the monitors of all the other intended users. Chatting is one of the most popular uses of the Internet. Generally, the users remain anonymous by using nicknames or pseudonyms to identify themselves online.

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

Internet chat works on client-server model, i.e., it requires a chat server and a chat client, the former usually provided by an Internet Service Provider (ISP), the latter is used by the persons chatting. Usually, the 'chat' 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.

A variation of chat is the phenomenon of the instant messenger. Instant messaging can be used to communicate privately with friends, relatives or co-workers. To send and receive instant messages, a user needs a connection to the Internet and instant messaging software such as, AOL Messenger, Yahoo Messenger and MSN Messenger. The instant messaging software enables a user to set up a list of contacts who also use the same program. Once this list is set up, one can see each contact as they come online. A chat session can be started with them quickly and easily. Most conversations are typed text messages that are sent back and forth, though more advanced users can exchange voice, video, files and more.

Conferencing

One of the most exciting features of Internet is to communicate, talk, and see 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 programs, 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 or use the audio alone, or just use the screen-sharing capability without either audio or video.

Audio Conferencing or Internet Telephony

Internet telephony also called IP telephony is a combination of hardware and software that allows the Internet to be used as a telephone carrier. Internet telephony is the conversion of analogue speech signals used on current telephone systems into digital data, allowing calls to be sent over the Internet, bypassing long distance charges. While the Internet was first devised as a way of transmitting data, it is now being used to make voice calls. Internet telephony is projected to explode as the costs plummet. After the costs of initial set-up and access to an Internet Service Provider, long-distance voice calls can be made

via the Internet free of charge, but current quality of voice transmission over the Internet is not always as good as direct telephone service. The following components are involved in making a phone call on the Internet:

- i) On the client side, a multimedia-equipped PC with special client software facilitates digitisation of voice. This can be done with a voice modem or other voice encoding method;
- ii) A direct or dial-up connection to the Internet allows the voice to be transmitted as packets to its destination;
- iii) Connection with the far side is achieved by IP address search, common servers or beacons to identify the called party (and to 'ring' that person's phone);
- iv) A similar arrangement on the far end completes the call and allows both parties to speak.

There are also PSTN / Internet gateways that allow regular telephone callers to make Phone-to-Internet-to-Phone connections. There are PC-to-Phone connections and Phone-to-PC connections.

A growing number of firms now sell Internet telephone handsets-devices that look like telephone handsets but plug into computer and let you dial, for free, anyone else who owns a similar handset. Large corporations are already adopting these VOIP systems.

Video Conferencing

Video conferencing is one of the most exciting areas of development in telecommunications, with applications ranging from business to government to education to home and family. Video conferencing 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, one can 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 of 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, video conferencing requires sufficient bandwidth to transfer video files at acceptable quality. Desktop conferencing, therefore, is not yet widely implemented for business and educational use. As bandwidth increases, desktop video conferencing is expected to blossom. The PictureTel and Vtel are two of the largest companies that sell video conferencing equipment.

NetMeeting

Microsoft NetMeeting facilitates a new way of talking, meeting, working and sharing over the Internet. It uses Internet phone voice communications and conferencing standards to provide multi-user applications and data sharing over Intranets or the Internet. Two or more users can work together and collaborate in real time using application sharing, whiteboard, and chat functionality.

NetMeeting can be used for common collaborative activities such as, virtual meetings. It can also be used for customer service applications, telecommuting, distance learning, and technical support. The product is based on ITU (International Telecommunication Union) standards; so, it is compatible with other products based on the same standards. Some of NetMeeting's built-in features are listed below.

- Place calls to anyone using the Internet or an Intranet
- Talking to someone over the Internet or an Intranet
- Seeing the person being called
- Working with others in an application
- Using the whiteboard to sketch in an online meeting

- Checking Speed Dial list to see contacts that are logged on
- Sending typed messages in Chat
- Sending files to everyone in a meeting

This software product supports text chat, video shared whiteboard, transferring of files, directory of connected users.

Using the directory, one can find someone to communicate with and call them. Communication can be established using text chat, video or audio provided both the parties have the necessary hardware support.

Files, such as, documents or pictures, can be exchanged. One of the problems with net meeting is the break-up in audio that sometimes it becomes inaudible when using it on the Internet, though it works fine over a high-speed company network.

NetMeeting is normally installed with Internet Explorer, or it can be download from <http://www.microsoft.com/windows/netmeeting/>. There is a website dedicated to NetMeeting at <http://www.netmeet.net/>.

NetShow

NetShow is basically a low-bandwidth alternative to video conferencing. It provides live multicast audio, file transfer and on-demand streamed audio, illustrated audio, and video. It is also a development platform on which software developers can create add-on products. According to Microsoft, NetShow takes advantage of important Internet and network communication technologies to minimise traffic while providing useful tools for multi-user collaboration. NetShow also uses streaming technology (as discussed above), which allows users to see or hear information as it arrives, rather than wait for it to be completely transferred.

Listserv

E-mail also provides a platform for sharing information quickly, and on a large scale by using listservs. Listservs are electronic groups that typically centre around broad topics 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 are potentially hundreds or thousands of people. It does not cost anything to subscribe to a Listserv, but simply requires that the user sends an e-mail message to the appropriate address with the message: subscribe (listserv) Firstname Lastname. Each Listserv 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.

News Groups

Another Internet service similar to Listserv is a News Groups. News Groups are like an International bulletin board. Each group is a forum for a different subject, where a subscriber can post his/her questions or answers. There are thousands of groups covering just about every area of interest. The difference between Listserv and News Group is that when a user joins a group, the mail is no longer automatically deposited into his/her mailbox. Instead a user is expected to go to the newsgroup himself/herself to read it. Some listservs can also be accessed as a newsgroup. A good analogy to a newsgroup is a bulletin board, i.e., one can go to it, as opposed to having mail delivered to his/her desk. The mails from newsgroup do not get cluttered and they can be easily regulated as to how often messages are read. A drawback of a newsgroup lies in the fact that a user must remember to go out to the newsgroup to look for information.

There are hundreds of newsgroup communities. They centre around topics such as, computing, news, recreation, social, and 'alternative' topics. Each newsgroup's 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 (and new groups appear every day). The only problem one might encounter is that it is left up to the service provider to determine which newsgroups will be made available on the new system. (this is comparable to the problem encountered with basic cable television).

Usenet

Usenet is a worldwide bulletin board system that can be accessed through the Internet or through many online services. The USENET contains more than 14,000 forums, called newsgroups, which cover every imaginable interest group. 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 program designed to access Usenet newsgroups), one can read an entire thread, i.e., 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, one requires a news viewer and access to an NNTP server. Most Internet Service Providers have such a server.

14.7.2 Connectivity

The Internet facilitates connection to servers on the Internet from remote locations. The services that can be used for establishing connection to servers on remote locations include the following:

Remote Login / Telnet

Remote Login is the ability of a computer user in one location to establish an online connection with another computer elsewhere. Once a connection is established with a remote computer, the user can use that remote system as if his/ her computer were a hard-wired terminal of that system. Within the TCP/IP protocol suite, this facility is called 'Telnet'. 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. 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 is used commonly to connect to electronic catalogues and databases of near and distant libraries.

Utilising Telnet, an Internet user can establish connections with a multitude of bibliographic databases (mostly library catalogues), campus information systems of various universities, full-text databases, factual databases and other online services. Since most of the databases now have a web interface, remote login / telnet is not used as often for accessing the databases.

Telnet, like most other services, is based on the client / server model. The client program, running on a PC, initiates the connection with a server program, running on a remote machine. Key strokes are passed from user's terminal directly to the remote computer

just as though they were being typed at a terminal of the remote computer. Output from the remote computer is sent back and displayed on user's terminal. Hyper Terminal is a client-end application built into Windows operating system that helps a user to access various servers through Telnet.

File Transfer Protocol (FTP)

File transfer is one of the most frequently used Internet applications, enabling a user to copy files from over a thousand different archives around the world. Archives stored on FTP server can be equated to repositories housing digitised information. 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 programs, graphic images, sounds, files formatted for particular software programs (e.g., files with word processing formatting instructions), and others.

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 authorisation that prohibits users from accessing files unless a user have a valid login name and password. Many other sites allow anonymous FTP, which provides unrestricted access to public files. Anonymous FTP access means that a user is not required to register himself to connect to the remote host. A user can connect to an anonymous FTP server using anonymous as login name and e-mail address as a 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).

Like most of the Internet applications, the FTP is also based on a client/server model. A user can run a program on his local machine called an FTP client, which in turn connects to another program running on a remote FTP server.

The following details are required for downloading a file from a remote server:

- a) The remote computer's Internet address (either the domain name address or the IP address);
- b) The pathname that tells the path of directories to get from the root directory to the directory where the file resides; and
- c) The file name itself.

14.7.3 Access to Information Resources

Gopher

A system; developed at the University of Minnesota, that pre-dates the World Wide Web for organising and displaying files on Internet servers. A Gopher server presented its contents as a hierarchically structured list of files. 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 follow a client-server model." With the ascendance of the web, many Gopher databases were converted to websites which can be more easily accessed via web search engines.

From about 1992 to 1996, Gopher was an Internet application in which hierarchically-organised text files could be brought from servers all over the world to a viewer on his / her computer. Gopher was a step towards the World Wide Web's Hypertext Transfer Protocol (HTTP), which effectively replaced the prior within a short time. With hypertext links, the Hypertext Mark-up Language (HTML), and the arrival of a graphical browser, Mosaic, the web quickly transcended Gopher. Two tools for searching Gopher file hierarchies were Veronica and Jughead.

Although the root directory for Gopher is sometimes still accessible at the University of Minnesota or elsewhere and despite some efforts to revive Gopher, virtually all Gopher servers are no longer active. It seems likely that almost all of the original Gopher content has been made accessible on the World Wide Web.

WWW as Information Resource

The web has established itself as the most widely used information resource by all kinds of people for variety of reasons. There are all kinds and types of information resources available on the web. Electronic resources on the Internet manifest themselves in numerous flavours and categories, although most of them emulate the traditional publishing while others are revolutionary in their design and approach. While the present trend to imitate and emulate the traditional models of scholarly communication may continue for some time, eventually the capabilities added by the new media would be used in more innovative ways. The information resources available via the web includes electronic journals, courseware, tutorials, manuals, patents, preprints, news, software, technical reports, theses and dissertations, bibliographic databases, images, audio, video, equipment / product catalogues, scientific data sets, library catalogues, museums and archives, virtual libraries, electronic books, online bookselling and print-on-demand, reference sources including dictionaries, encyclopaedias, biographies, abbreviations, thesauri, subject headings, handbooks, maps, organisations and people including employment and career sources, funding / grants sources, libraries / information centres, organisations, research institutes, companies, societies, people, experts, scientists, directories and subject portals.

14.7.4 Searching Information Resources on Internet

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 12,000 servers and over 2 million files. The Archie service is accessible through an interactive elnet 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-date 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 the 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.

Veronica

Veronica was a program that allowed a user to search the files of the Internet's Gopher servers for a particular search string. Veronica indexed the world of Gopher servers, much like Archie searches all anonymous FTP servers. Veronica was an indexing spider that visited the Gopher sites, read all the directory and file names, and then indexed them in one large index. However, with the almost complete demise of Gopher servers, Veronica has become a relic of the early 1990s.

Wide Area Information Server (WAIS)

Developed by Thinking Machines in collaboration with Apple Computer, the Wide Area Information Server (WAIS) is an Internet system in which specialised subject databases are created at multiple servers distributed at various locations and made accessible in real-time to the users with WAIS client programs. A user can either download a WAIS client or use Telnet to connect to a public WAIS client. The WAIS is used as a back-end search engine in combination with the web. WAIS uses its own Internet protocol, an extension of the Z39.50 standard. A directory of the servers keep track of servers distributed at various locations. The user of WAIS is provided with or obtains a list of distributed databases. The user enters a search argument for a selected database and the client then accesses all the servers on which the database is distributed. The results provide a description of each text that meets the search requirements. The user can then retrieve the full text. WAIS relies on indexed data collections, or libraries. These libraries are the collections that consist mostly of informational material.

Most web users would consider WAIS as superfluous with abundance of server files and search engines. However, librarians and specialist users may find some specialised information available through WAIS that is not currently available on the web.

The WAIS was developed as a generalised retrieval system for accessing data around the world i.e., to provide a common interface to 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 search engines support Boolean queries, truncation and proximity searching.

Self Check Exercise

- 9) What is the difference between WAIS and a search engine?
- 10) What are the major differences between listserv and newsgroup?

- Note :** i) Write your answers in the space given below.
ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

14.8 INTERNET FOR LIBRARY APPLICATIONS

The Internet has made a great impact on almost all aspects of a library. It has affected the work processes, services, collection development processes, types of collection, users' instructions, readers' services and preservation of the intellectual records. Internet can potentially support a range of traditional and non-traditional library services. Most of the library services generated using digital resources resemble closely those generated manually with improvements and modifications to suit the requirements of automated services. However, Internet has also been used to generate innovative services that did not have a counterpart in manual practices. Some of the important applications of Internet for supporting traditional library activities and for Internet-based library services are discussed in the following sub-sections:

Acquisition and Collection Development

Amazon.com started a new phenomenon on the web with its online bookshop, which has been expanded to include other products like CDs, music, electronics, toys, art works, computers, and other store items. Amazon.com was termed as the "Earth's Biggest Library" (<http://www.infotoday.com/newsbreaks/nb1122-1.htm>) although it is not performing all the functions of a library. There are several sites that are now in the business of online book selling. Most book publishers and bookstores announce their new books through their websites. The libraries can use these sites as a source for acquiring information about new books for collection development and for verifying information on various aspects including costs, etc. Some of the important sites that are in the business of online book selling are as follows:

Amazon.com Bookstore	http://www.amazon.com/
Barnes & Noble	http://shop.barnesandnoble.com/
Best Book Buys	http://www.bestbookbuys.com/
Book Finder	http://www.bookfinder.com/
Catalog Site	http://www.catalogsite.com/
Pricescan Before You Buy	http://www.pricescan.com
Studentstbookworld.com	http://www.studentsworld.com/
Swotbooks.com	http://swotbooks.com/
Varsitybooks.com	http://www.varsitybooks.com/

Besides using Internet for selection of books in printed format, the libraries can also buy access to various types of electronic resources available on the Internet. The information resources available on the Internet can be taken into account in the process of collection development. The libraries of future will not become digital libraries, but will rather acquire access to ever growing digital collections on behalf of their users. Majority of these collections are being made available by external sources like commercial publishers, collections mounted by scholarly societies, resources at other libraries, electronic journal sites, etc. The electronic journals have become the largest and fastest growing segment of digital collections for most libraries. There are more than 20,000 peer-reviewed electronic journals that are now available in electronic form either exclusively or in addition to their print versions.

Besides electronic journals, electronic books, encyclopaedia, dictionaries, directories, online databases, online courseware, etc. are also available on subscription / purchase model.

Technical Processing of Books

The Internet can be used to access latest authoritative tools related to classification and cataloguing to render help in better and faster processing of documents. Links to tools like Library of Congress Classification Schedule, LC Subject Headings, MARC Documentation, OCLC User Documentation, other thesauri and subject dictionaries available on the web can render required help to the staff members involved in technical processing of books. Also, the library can provide a monthly / weekly list of books acquired by the library through the library website.

Journals Ordering and Management

Most publishers maintain their websites with subscription information about their journals that can help the staff involved in journal selection, ordering and management. Several

websites maintain meta resources with links to journals from various publishers. The library can provide a complete list of journals subscribed by it as well as weekly list of issues of journals received by it. List of complete journals holding of the library on the website would also enhance the usage. Link to union catalogue of journals subscribed by a group of libraries would help in resource sharing.

Reference Services

The number of reference sources available on the web has increased exponentially during past decade. The libraries can develop subject portals that would provide links to important reference sources through the Library website. Meta sites like Xrefer provides a basket full of reference sources with a single search interface to help the users to retrieve information in a quick and efficient way.

14.8.2 Traditional Library Services Modified in the Internet Era

OPAC to WebPAC

Remote access to the library catalogues was possible only through a Telnet connection till recently. The web-based interfaces are now available with most of the integrated library software packages including Libsys. Websites are increasingly providing links to their WebPAC instead of Telnet links to Their Library OPAC. Exploiting the provisions of hyperlinking that the web provides, various searchable elements of a bibliographic record in a WebPAC are hyperlinks to other records in the database. For example, an author is a hyperlink to all records in the database for that author, a series is an hyperlink to all serial titles under that series; a keyword for a record is a link to all records in database having that keyword, and so on. In effect, a WebPAC adds software-based functionality to a conventional OPAC. A user thus has additional incentives to visit the library web page hosting WebPAC. With web-based resources and services in place, many libraries are phasing out their dumb terminals. The library websites are increasingly becoming a more logical gateway to the catalogue and other web-based library resources. The acceptability of web-based interfaces to the Library OPAC is much greater because web interfaces are familiar to the users with its graphical and navigational interfaces. The users can click complex subjects instead of typing them in or remembering complex unix commands.

CD-ROM to Web-based Indexes and Databases

Availability of CD-ROM in the late 1980s, as a medium with high storage capacity, longevity, and ease of transportation triggered production of several CD-ROM information products which were earlier available through online vendors or as conventional abstracting and indexing services in printed format. Some of the important secondary services including "Guide to Current Periodical Literature" (H.W. Wilson) discontinued their print version in favour of CD-ROM version, which had improved functionality in terms of search and browsing interfaces. The libraries are witnessing yet another migration from bibliographic databases on CD-ROM to web-based bibliographic databases akin to the one that was witnessed earlier from print-based secondary services to CD-ROM databases. This phenomenon has further been fuelled with availability of web-interfaces for most of the online search services. The web-based interfaces provide several advantages to users that are either not possible or not yet available on CD-ROM. Most web-based bibliographic databases use hyperlinks and other facilities possible in web documents including links to the full-texts of articles available on a publisher's website. Several bibliographic databases have discontinued their CD-ROM versions in favour of web-based versions. Besides the advantages mentioned above, migration to web-based services opens up resources to remote users.

From Manual Reference Service to Digital Reference Service

Reference service and providing advice to the library users on the information sources are key areas of information service for any library. The technology now allows reference

librarians to reach out to the users using the network instead of waiting at the reference desk for users to come by. Besides, advising users on the mechanisms of using the library, a reference librarian is also involved in delivering reference service that requires deep intellectual understanding of subject. Although automated libraries are not yet sufficiently advanced to offer interactive reference services, electronically mediated reference services are increasingly available through libraries and information centres.

Digital reference service, also called 'Ask-An-Expert' or 'Ask-A-Librarian' service is an Internet-based question and answer service that connects users with individuals who possess specialised subject knowledge and skill in conducting precision searches. As opposed to static web pages, digital reference services use the Internet to place people in contact with people who can answer specific questions and instruct users on developing certain skills. The people who serve as digital reference experts (also called volunteers or mentors) are most of the time information specialists, affiliated to various libraries.

How does it Work?

Most 'Ask-a-Librarian' services have a web-based question submission form or an e-mail address or both. Users may submit questions by using either of them. Once a question is read by a service, it is assigned to an individual expert for answering. An expert responds to the question with factual information and or a list of information resources. The response is either sent to the user's e-mail account or is posted on the web so that the user can access it after a certain period of time. Many services have informative websites that include archives of questions and answers and a set of FAQs. Users are usually encouraged to browse archives and FAQs before submitting a question.

Virtual Reference Desk (<http://www.vrd.org/>) provides resources and links to experts that offer digital reference services. The site hosts searchable database of high quality 'ASK-A' service along with alphabetical and subject wise listing. Virtual Reference Desk also hosts a listserv called 'Dig-Ref' to promote and explore the growing area of digital reference services.

From Manual Reference Service to Real Time Digital Reference Service: Library Chat Rooms

Several libraries have started experimenting with offering real time digital reference service, using chat software, call counter management software, web contact software, and providing interactive information service, bulletin board services, interactive customer assistance, etc.

Many libraries are experimenting with Internet chat technology as an innovative method to extend and enhance traditional and remote reference service. While digital reference service is an asynchronous method of information delivery, the Internet chat provides the benefit of synchronous communication between a user and a reference librarian (or mentor). Interactive reference services facilitate a user to talk to a reference librarian at any time of day or night from any where in the world. Unlike with e-mail reference, the librarian can perform a reference interview of a sort by seeking clarifications from the user. The librarian can conduct Internet searches and push websites onto the patron's browser, and can receive immediate feedback from the patron as to whether his or her question has been answered to satisfaction. Most libraries currently involved in real-time reference service are part of a collaborative network so that they can share staffing and work round the clock to provide highly efficient reference service any time. Library of Congress Collaborative Digital Reference Service is one of such services. Several institutions including Cornell University, Internet Public Library, Michigan State University, North Carolina University are offering Internet chat-based service using software like LivePerson, AOL Instant Messenger, Conference Room and Netscape Chat. The librarians have observed that their relatively new chat-based service logged significantly more questions in a relatively short time than did their well established e-mail digital reference service.

LiveRef(sm) (<http://www.public.iastate.edu/~CYBERSTACKS/LiveRef.htm>) maintains an online registry of real-time digital reference services.

From Manual Document Delivery to Electronic Delivery Services

Abstracting and indexing services have proved themselves as most effective means of finding recent and retrospective published research work. The effectiveness of these secondary services further enhanced with availability of these secondary services on CD-ROM with efficient search interfaces and other features that are possible only in electronic media.

Once a researcher gets bibliographic references relevant to his research work, the more arduous task of locating the full-texts of research article begins. While the parent library may cater only to 10 - 20 % of his references, remaining articles may have to be arranged through Inter Library Loan (ILL) or through Document Delivery Services (DDS) which can be very time consuming. Most libraries have been using commercial (Informatics India) and non-commercial (BLLD and NISCAIR) document delivery services to ensure quick and efficient access to primary information for the library users. Most online search services like DIALOG, ESA/IRS and STN were offering manual document supply services since their inception. The process was labour-intensive and time consuming.

The term 'Electronic Document Delivery Systems' (EDDS) implies delivery of electronic version of a document that might involve reproduction of an electronic copy of a document if it is not already available in electronic format. The libraries had been using fax machines for immediately delivering photocopies of articles via telephone lines. The first use of electronic document delivery was based on scanning technology. With maturity of scanning equipment and technology, document supply services started scanning the documents as bitmap page images. Applications are built in such a way so as to automatically produce a hard copy together with a header page containing the address of the applicant, which can again be sent, by snail mail or facsimile. A software package known as 'Ariel' is used in several libraries in the developed countries for delivery of scanned articles via the Internet. The Ariel software, loaded on an Internet-enabled computer, can receive and send electronic information to other libraries which have installed Ariel. The ADONIS system developed in the late 1980s is a document delivery system based on bit-mapped page images.

Availability of most of the peer reviewed research journals in electronic format, inexpensive technology to scan articles and improved electronic delivery mechanisms are some of the enabling factors that have contributed to well-established electronic document delivery systems now available commercially. More recently, most of the secondary services that were available on CD-ROM or through online search services are now available on the Internet where the journals are linked to the publisher's site. The technology has now been perfected and there are several electronic document delivery services that allow a user to download an article in full-text from their site or deliver them electronically as attachment to e-mails. Most electronic publishers and aggregators like OCLC, Blackwell, OVID, etc. are offering full-text of articles through their websites. Different vendors have various payment options; some charge each time the journal is used, whereas others provide open access for a set annual fee. A user who wishes to have the item delivered can enter a credit card number and specify a delivery method (postal, UPS, fax, e mail, etc.) and indicate whether it is a rush item (with a rush order fee attached).

Some of the important Electronic Document Delivery Services include:

Uncover	http://uncweb.carl.org/uncover/subtitle.html
Articles in Physics	http://ojps.aip.org/
Bioline Publications	http://bioline.bdt.org.br/journals
Chemport	http://www.chemport.org/
ScienceDirect	http://www.scienceDirect.com/
OCLC	www.oclc.org/
Northern Light	www.northernlight.com/

14.8.3 Internet-based New Library Services

Virtual Library Tours

Several library websites facilitate virtual guide to the physical facilities including collections, services and infrastructure available in the library through their websites. The combination of the following three web-based interfaces are used to facilitate the virtual library tours:

Library Maps and Floor Plans

Most library websites provide library layouts and floor plans to guide users to physical location of facilities and services along with links to relevant information. Client-side image maps are used to make various parts of floor plans as clickable image maps. An example can be seen at the Home Page of the Central Library, IIT Delhi at: <http://www.iitd.ac.in/acad/library/layout.html>

Library Websites

Academic libraries in the developed countries started using web technology to create home pages as starting points or as gateways for searching information about the library. A home page reflects the characteristics of an academic institution. It provides an opportunity to the library to propagate its services and facilities to the academic community worldwide. The home pages of libraries are increasingly used as an integrated interface designed to deliver detailed information about a library as well as to provide access to all its computer-based services.

Besides offering information, the library websites of academic institutions invariably host subject gateways or subject portals that contain links to web resources for subjects of interest to the institution. Most of the services (modified or new) included in this section are offered through the websites of most of the academic institutions especially in the developed world. The IIT Delhi Central Library also offers several of these services through its website available at <http://www.iitd.ac.in/library/>.

Subject Gateways or Library Portals

The web, being a hypermedia-based system, allows linking amongst electronic resources stored on servers dispersed geographically on distant locations. The portal sites or gateways redirect a user to the holders of the original digital material. The librarians, being among the earliest users of the web, and following their professional instincts, immediately began to create links to collections on all sorts of subjects.

A subject gateway can be defined as facility that allows easier access to web-based resources in a defined subject area. The simplest type of subject gateways is sets of web pages containing lists of links to resources. Some gateways index their lists of links and provide a simple search facility. More advanced gateways offer a much-enhanced service via a system consisting of a resource database and various indexes, which can be searched and / or browse throughout a web-based interface (O'Leary, M., 2000).

Subject gateways are also known as Subject-based Information Gateways (SBIGs), subject-based gateways, subject index gateways, virtual libraries, clearing houses, subject trees, pathfinders and other variants thereof. Subject gateway is an important component of a library website designed for the library users so as to help them discover high-quality information on the Internet in a quick and effective way.

In the traditional information environment human intermediaries, such as, publishers and librarians, filter and process information so that users can search catalogues and indexes of organised knowledge as opposed to raw data and information. Subject gateways work on the same principle, i.e., they employ subject experts and information professionals to select, classify and catalogue Internet resources to aid search and retrieval for the users. Users are offered access to a database of Internet resources descriptions which they can search by keywords or browse by subject area. A description of each resource is provided to help users assess very quickly its origin, content and nature, enabling them to decide if

it is worth pursuing them further. In the process users get benefited from the expertise of librarians and subject experts with subject gateways rather than having to locate, evaluate, filter and organise the resources themselves. Specialised software are available as freeware or as priced software to create and maintain professionally developed subject gateways. Some of the major portal sites or gateways that provide access to electronic resources on the Internet are as follows:

WWW Virtual Library	http://www.edoc.com/
Internet Public Library	http://www.ipl.org/
Michigan Electronic Library	http://mel.lib.mi.us/
Penn Electronic Library	http://www.library.upenn.edu/resources/
BUBL Information Service	http://bubl.ac.uk/
Argus Clearing House	http://www.clearinghouse.net/
Internet Index	http://sunsite.berkeley.edu/InternetIndex/

Web-based User Education

The WWW provides a dynamic environment for distributing information over a large network and web-based instruction is a suitable applications of this facility. Web-based guides and teaching tools can be easily updated, accessed, and printed on demand. They may include colour graphics, and screenshots. The web-based user education provides a high degree of interactivity and flexibility to the users offering them the benefit of self-space, graduated to teach from basic to highly advanced levels and designed in a wide range of formats that accommodate diverse learning styles. The proliferation of digital resources will generate greater demands on reference and instructional services. With availability of digital resources that can be used any where at any time, requirement for instructional and reference services will also grow. Failure to develop both the technological aspects and required service components would lead to under utilisation of digital resources. The library websites can use web-based user education for imparting training to users in the following areas:

- i) Basic library skills along with glossary of library terms;
- ii) Using Library OPAC / WebPAC, locating books, magazines and other library materials;
- iii) Searching CD-ROM and web-based databases and other electronic resources; and
- iv) Subject searching, using Boolean operators and searching Internet resources through search engines.

The web technology provides for incorporating both synchronous and asynchronous interaction in the web-based user education.

Frequently Asked Questions (FAQ)

Most library websites have Frequently Asked Questions (FAQ) along with their answers. Some libraries have database-driven FAQs along with search interface. These FAQs are generally on the services and facilities that the library provides. These FAQs generally do not include reference questions.

Library Calendar

The library calendar lists events or show information for forthcoming events. Library calendars can have improved look and functionality with JavaScript or special software.

Most library websites have web forms for inviting feedback from the users on different aspects of the library. These include the following:

- i) Suggestions for services;
- ii) Users Survey;
- iii) Comments on the website and suggestions to improve it;
- iv) Requests to library to acquire certain titles or materials;
- v) Reference Service (often Ask-a-Librarian); and
- vi) Inter-library loan or other document delivering services.

Bulletin Boards, Threaded Discussion Forums and Listservs

Several libraries are using bulletin boards, threaded discussion forums and listservs to help promote and evolve web-based library services. Most libraries use bulletin board system as an electronic message system to propagate or announce the services and new activities. The bulletin board system is also used as an interactive interface to invite suggestions on activities and services of the library. It can also be used as an interface to provide library services. Messages in a bulletin board system can be written by any one and are stored in a common area for any one to read.

Discussion forums allow open exchange of messages on a topic of common interest. They allow users to initiate a discussion on new topic, replying to an ongoing discussion (called thread). All messages for a given topic or thread are grouped together for the convenience of users. Discussion forums are basically modified bulletin boards, which have added feature of dividing messages into logical groupings called thread. Threads enable a person to focus on a particular topic and see input from many individuals making comments on the topic.

A listserv is a mailing list program wherein a group of people with common interest is connected by e-mail. Any mail sent to the listserv is distributed to all those who have signed up for the list. Several libraries host listservs for the users for providing them a platform to discuss and share their views on books that they have read, or discuss specific books / authors and so on.

14.9 SUMMARY

The Internet holds tremendous potential for whole array of activities including online distance education, global digital library, e-commerce, Internet telephony, electronic publishing, electronic journals, virtual museums, etc. It has particularly established itself as a powerful medium for self-education for people in isolated or remote areas for its ease-to-use, familiarity with masses, availability of tools and wider accessibility. This Unit elaborates on various Internet services and its potential applications.

The Unit describes the World Wide Web (WWW) as the most successful networked multimedia hypertext-based distributed system of our time. It elaborates on the capacity of the Web as an interface to all other services offered by the Internet including ftp, telnet, Gopher, etc. The Unit describes how the web works, its client-server architecture, Hypertext Transfer Protocol (http), URL and the web servers. Various types of web browsers, their functionalities are described in detail. The Unit deals with mark-up languages including SGML, XML, HTML, dynamic HTML and VRML. The Unit also elaborates on various HTML tags, style sheets and layering.

by the Internet. These are broadly grouped into the following four categories:

- i) Internet-based communication services
 - a) Communication amongst individuals: electronic mail, Internet telephony, Internet chat,
 - b) Communication services for groups: Internet relay chat, electronic discussion groups, Listserv, News groups, Usenet and video conferencing.
- ii) Connectivity: Telnet, remote login and file transfer (FTP)
- iii) Information resources: Gopher, WWW, FTP servers
- iv) Searching information resources on Internet: WAIS, search engines and web directories.

The Unit describes each of the above-mentioned Internet services in detail. Lastly, the Unit elaborates on applications of Internet for supporting traditional libraries activities and for Internet-based library services. A range of traditional and non-traditional library services that the Internet supports are also described. Most of the Internet-based library services resemble closely those generated manually with improvements and modifications to suit the requirements of automated services. However, Internet has also been used to generate innovative services that did not have a counterpart in manual services. The Unit deals with both types of Internet-based library services.

14.10 ANSWERS TO SELF CHECK EXERCISES

- 1) The World Wide Web is a distributed system of graphical, hyper-linked information, based on the hypertext transfer protocol (HTTP). The World Wide Web is also known as WWW or W3. It is a global, networked system of dedicated host computers that serve documents (files) formatted in HTML. It is called multimedia hypertext system because web pages can contain text, images and multimedia components, can include hyperlinks to other such documents on different servers, and can also act as interface, linking users with underlying special-function applications. The web debuted in 1993. The web was originally conceived as a platform-independent tool that scientists could use to exchange documents about their work.
- 2) The Client-server architecture distributes the processing of a computer application between two computers, the client and the server. The client is normally a PC. The client application program accesses data and performs processing on the server and using the data obtained via the server, processes more tasks at the client. As computers became more powerful and readily available, the client-server model became popular.
- 3) The Hypertext Transfer Protocol (HTTP) is the set of rules for exchanging files on the web. The concept of HTTP is that files can contain links or references to other files whose selection would lead to transfer of requests from one file to another. Any web server machine contains, in addition to the HTML and other files it can serve, an HTTP daemon, a program that is designed to wait for HTTP requests and handle them when they arrive. The web browser is an HTTP client, sending requests to server machines. When a user requests for a file through browser by either 'opening' a web file (typing in a Uniform Resource Locator) or clicking on a hypertext link, the browser builds an HTTP request and sends it to the Internet Protocol address indicated by the URL. The HTTP daemon in the destination server machine receives the request and, after any necessary processing, the requested file is returned.
- 4) The Uniform Resource Locator is the distinct address that identifies each resource on the Internet. More formally, an URL is the networked extension of the standard

- 4) The Uniform Resource Locator is the distinct address that identifies each resource on the Internet. More formally, an URL is the networked extension of the standard filename concept that can exist on any machine on the network and served via any of several different methods. An example of a URL is:

<http://www.iitd.ac.in/acad/library/index.html>

The URL above describes a web page that can be accessed with an HTTP (web browser) application that is located on a computer named www.iitd.ac.in. The specific file is in the directory named /acad and subdirectory /library and is the default page in that directory (which, on this computer, is named as index.html).

- 5) A web browser is a software on client's machine that allows a user to browse the Internet. There are two important web browsers, i.e., Internet explorer and Netscape Navigator. Web browsers can read a number of different protocols, specified in the URL. In the case of http, the web browser translates HTML coding into what you see on the screen: a combination of text, images, links, etc. The browser also allows to navigate from link to link, traversing the Internet or exploring one site after the other with each mouse click. Internet explorer is most widely used web browser.
- 6) Plug-ins or helper applications are external software programs that allow web users to view or hear multimedia presentations, regardless of platform. Plug-ins can easily be installed and used as part of a web browser. Plug-ins or helper applications extend and enhance the capabilities of web browsers and are needed to handle many of the newer hypermedia such as, streaming audio, vector graphics, three-dimensional multimedia and virtual worlds. Browsers hand over data to appropriate helper applications such as, RealAudio, Adobe Acrobat, QuickTime, Shockwave and others.
- 7) The mark-ups are a set of file markers, field tags, and / or hypertext links in the text. Mark-up can be based on a standard mark-up language, such as, the Standard Generalized Mark-up Language (SGML). The particular file marker and field tags may be predefined, as in case of HTML or they can be declared while setting up a text indexer for a file or a document. Hypertext links do not have to be declared, because they are automatically indexed as searchable terms.
- 8) Some of the major differences between HTML and XML are:

HTML deals with the appearance of a document whereas XML deals with its structure.

Hierarchical element structure: Unlike HTML, the XML documents must have a strictly hierarchical tag structure. That is, start tags must have corresponding end tags. In XML vocabulary, a pair of start and end tags is called an element. Any element must be properly nested within another.

Empty tags: Empty tags are also allowed as elements in XML documents. An empty tag is essentially a start and end tag in one, and is identified by a trailing slash after the tag name.

Declared entities: All entities must be declared in a DTD. XML entities are analogous to constants in other languages. Entities can be expanded during processing, like a macro-preprocessing capability, saving error-prone duplication of common text. We won't cover entities any further, though, since we don't use them in our example. Entities are an important topic, so you may want to

Tags / Mark-up: HTML has pre-defined sets of mark-ups. XML facilitates creation of new tags required for data to be coded. The new tags, should, however, be defined in a document called Document Type Declaration (DTD).

Case sensitivity: Unlike HTML, the XML tags are case-sensitive.

Relevant white space: Unlike in HTML, the white space in the data between tags is relevant in XML because XML is a data format. However, within the markup itself, and also within quoted attribute values, white space is normalized, or removed.

Character encoding: XML allows a user to specify different character set encodings. The encoding must be identified within Document Type Declaration.

- 9) The WAIS was 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 on a client-server model and has the following 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.
- 10) A user can either download a WAIS client or use Telnet to connect to a public WAIS client. The WAIS is used as a back-end search engine in combination with the web. WAIS uses its own Internet protocol, an extension of the Z39.50 standard. While the WAIS brings results on the web in real time, the search engine searches for indexing terms residing on the databases of search engines collected by their spiders.
- 11) The difference between listserv and news group is that when a user joins a group, the mail is no longer automatically deposited into his mailbox. Instead, the user is expected to go to the newsgroup himself to read it. A good analogy to a newsgroup is a bulletin board, i.e., one can go to it, as opposed to having mail delivered to his / her desk. The mails from newsgroup do not get cluttered and they can be easily regulated as to how often messages are read. A drawback of a newsgroup lies in the fact that a user must remember to go out to the newsgroup to look for information.

14.11 KEYWORDS

- Active X** : A set of platform independent technologies developed by Microsoft that enable software components to interact with one another in a networked environment, like the Internet. In addition to adding functionality within the browser (for example, by enabling Microsoft Word to be opened in a browser) Active X components can be exploited by malicious mobile code.
- Archie** : Archie is a software for locating files on anonymous FTP sites. Archie is an indexing spider that visits each anonymous FTP site, reads all the directory and file names and then indexes them in one large index. A user can then query Archie, which checks the query against the index.
- ASP** : Stands for Active Server Pages, a Microsoft's technology to enable HTML pages to be dynamic and interactive by embedding scripts, i.e., either VBScript or JScript, Microsoft's alternative of JavaScript. Since the scripts in ASP pages (suffix .asp) are processed by the server, any browser can work with ASP pages regardless of its support for the scripting language used therein.
- CAD** : Stands for Computer Aided (or Assisted) Design or Drafting, where in combinations of computer hardware and software is used for designing products or architecture. Sometimes CAD output is sent directly to operate the machinery that makes the product. This is called Computer Aided Manufacturing.
- CSS** : Stands for Cascading Style Sheets; when attached to documents. They describe how the document is displayed or printed, e.g., a CSS sheet is attached to an HTML document, to influence its layout when accessed

via a browser. CSS supports cascading, i.e., a single document may use two or more style sheets which are applied according to specified priorities.

- CGI** : Stands for Common Gateway Interface, CGI is a set of rules that describe how a web server communicates with another piece of software on the same machine, and how the other piece of software (the 'CGI program') talks to the web server. Any piece of software can be a CGI program if it handles input and output according to the CGI standard. Usually, a CGI program is a small program that takes data from a web server and does something with it, like putting the content of a form into an e-mail message, or turning the data into a database query.
- DHTML** : Dynamic HTML is simply the combined use of both CSS and JavaScript together in the same document.
- DTD** : Document Type Definition, is a collection of XML markup declarations that, as a collection, defines the legal structure, elements, and attributes that are available for use in a document that complies to the DTD.
- File Transfer** : File transfer is a common method of moving files between two Internet sites. FTP is a special way to login to another Internet site for the purposes of retrieving and / or sending files. There are many Internet sites that have established publicly accessible repositories of material that can be obtained using FTP, by logging in using the account name anonymous; thus these sites are called anonymous ftp servers.
- Gopher** : A method of making menus of material available over the Internet. Gopher is a Client and Server style program, which requires that the user have a Gopher Client program. Although Gopher spread rapidly across the globe in only a couple of years, most of the Gopher sites have been converted into websites.
- HTML** : Stands for Hyper Text Markup Language. It is the authoring language used in the creation of documents for the World Wide Web. HTML was initially created for use as a universal common document language for the World Wide Web. It indicates the type of information rather than the exact way it is to be presented. The actual presentation is left to the software that converts the contents to a suitable format for viewing. Text in an HTML document can be translated on-the-fly by a machine translator whereas text embedded in images and graphics (gifs, jpegs) must be localized.
- HTTP** : Stands for HyperText Transport Protocol, the protocol for moving hypertext files across the Internet. It requires an HTTP client program on one end, and an HTTP server program on the other end. HTTP is the most important protocol used in the World Wide Web.
- HTTDP** : Stands for Hyper-text Transfer Protocol Daemon, a computer program, which manages the transfer of hypertext and multimedia, documents over the Internet.
- Hypertext** : Any text that contains links to other documents - words or phrases in the document that can be chosen by a reader and which cause another document to be retrieved and displayed.
- IRC** : Short for Internet Relay Chat, IRC is a program that enables Internet users to trade messages in real time, simulating a conversation. Users can 'talk' to one another on the Internet using IRC technology. Communication occurs through predefined chat channels, which are

thematically arranged so that people discussing on these channels stay within a certain range of topics.

- Java** : Developed by Sun Microsystems, Java is a programming language that is specifically designed for writing programs that can be safely downloaded to a computer through the Internet and immediately run without fear of viruses or other harm to the computer or files. Using small Java programs (called 'Applets'), web pages can include functions such as, animations, calculators, and other fancy tricks. Java is a simple, robust, object-oriented, platform-independent multi-threaded, dynamic general-purpose programming environment. It is best for creating applets and applications for the Internet, intranets and any other complex, distributed networks.
- JavaScript** : JavaScript is a scripting language from Netscape. Java and JavaScript are not the same. While Java was built as a general-purpose object language, JavaScript is intended to provide a quicker and simpler language for enhancing web pages and servers. JavaScript is embedded as a small program in a web page that is interpreted and executed by the web client. JavaScript functions can be called from within a web document, often executed by mouse functions, buttons, or other actions from the user.
- Listserv** : A mailing list program for communicating with other people who have subscribed to the same list. Using e-mail, one can participate in listservs pertaining to any topics of interest. When a user submits a message to the server, the message is relayed to all those on the listserv. The messages are received from other participants via e-mail.
- ODBC** : Stands for Open Data Base Connectivity. ODBC is a de facto standard programming interface developed by Microsoft for accessing a database. Because it is a standard applied across DBMS and applications, communications between different platforms and DBMS is transparent using this interface.
- SGML** : Stands for Standard Generalized Markup Language. SGML is a standard for how to specify a document markup language or tag set. SGML is not in itself a document language, but a description of how to specify one. It is a meta language. HTML and XML are examples of SGML-based languages.
- Telnet** : It is a program that enables a user to open an interactive login session with a host machine over TCP/IP networks like the Internet. Commands can be entered from a PC and are executed on the server exactly as if the user is seated at the remote machine. Most computers require that a user should have an account for access.
- URL** : Stands for Uniform Resource Locator. It is the global address of a document or other resource on the World Wide Web. The first part of the address indicates what protocol to use, and the second part specifies the IP address or the domain name where the resource is located.
- Usenet** : Also known as 'news', 'newsgroups', or 'discussion groups'. Usenet is a huge collection of newsgroups about every and any topic one can imagine. Once a person posts to a newsgroup, it is transmitted to news servers all over the world for other people to see.
- VERONICA** : Stands for Very Easy Rodent-Oriented Net-wide Index to Computerized Archives. Developed at the University of Nevada, Veronica was a database of the names of almost every menu item on thousands of

Gopher servers. VERONICA has disappeared since most Gopher sites have been converted into websites.

- VRML** : Stands for Virtual Reality Modeling Language. VRML is an open, platform-independent file format for 3-D graphics on the web. It encodes computer-generated graphics in a way that makes them easily transportable across the network. VRML requires a special web browser to display these graphics, which simulate virtual reality 3-D 'environments', or 'worlds' through which the user can move and interact with objects. These 3-D 'worlds' can contain objects that link to documents, other objects, or other 3-D worlds.
- WAIS** : Stands for Wide Area Information Server. WAIS is a commercial software package that allows indexing of huge quantities of information and then making those indices searchable across networks such as, the Internet. A prominent feature of WAIS is that the search results are ranked (scored) according to how relevant the hits are to the search topic.
- Web Chat** : web chat is an online tool that enables users to communicate with each other through real-time text messaging without making a telephone call or waiting for an email response.
- World Wide Web** : The World Wide Web also called WWW, W3, or just the web is a collection of information servers linked together through hypertext using a language called HTML. The HTML, the de facto language for the web allows text embedded with graphics, audio and other file types. This allows a user to select a hypertext link on one page, which may take him / her to a different server located distantly, may be even halfway around the world. Users point and click to access World Wide Web pages using browser software, such as, Internet Explorer, Netscape Navigator, Lynx, etc. that provides the front-end once the Internet connection is actually established.
- Web Browser** : A software application used to make navigating the Internet easy for the user by providing a graphical user interface (or GUI) so that the user can click menus, icons, or buttons rather than learning difficult computer commands. Also called a web client because the browser application resides on the client, or the computer of the individual using it, rather than residing on a web server.
- Web Server** : A web server is either software that manages websites or the hardware on which server software is run. A server may be linked to the World Wide Web or it may be an 'internal only' server, meaning only certain individuals may have access to it. Web servers also regulate the use of CGI programs so that input and output may be made dynamic. A web server simply gets a browser request and sends the appropriate web page or data.
- XLS** : System of style sheets for use with XML.
- XML** : Stands for Extensible Markup Language (XML), XML is an extremely simple dialect of SGML. The goal is to enable generic SGML to be served, received, and processed on the web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

14.12 REFERENCES AND FURTHER READING

Computerscope. *The Worlds Leading Resource for Internet Trends and Statistics*. (http://www.nua.ie/surveys/how_many_online/world.html)

Dawson, A. (1997). *The Internet for Library and Information Professionals*. London: Library Association Publishing.

Dern, Daniel (1994). *The Internet Guide for New Users*. New York: McGraw Hill.

Ellsworth, [et al.]. (1997). *The Internet 1997*. Indianapolis: Sams.Net Publishing.

Falk, H. (1997). World Wide Web and Retrieval. *Electronic Library*. 15(1).

Hahn, Harley. (1997) *Internet: Complete Reference*. 2nd ed. New Delhi: Tata McGraw Hill.

Internet Economy Indicator (<http://www.internetindicators.com/factfigure.htm>)

Issues of Growth and Evolution in the Internet. (<http://www.zvon.org/tmrhc/rfc1380/output/chapter2.html>)

Johnson, Dave (1998). *Internet Explorer 4: Browsing and Beyond*. New Delhi: Tata McGraw Hill.

Kumar, PSG and Vashishth, CP.(1999).CALIBER - 99: Academic Libraries in Internet Era: Paper. *Sixth National Convention For Automation of Libraries in Education and Research, Nagpur*, 18 - 20 Feb. 1999. Ahmedabad: INFLIBNET.

Levine, John R., [et al.]. (2000). *The Internet for Dummies: Starter Kit*. 7th ed. New York: John Wiley.

Living Internet (<http://www.livinginternet.com/>)

Lowe, Doug (2001). *Internet Explorer 6 for Dummies*. New York: John Wiley, Inc.

McBride, P.K. (1999). *Internet Made Simple*. 2nd ed. Oxford: Butterworth-Heinemann.

Mehta, Subhash (1996). *Understanding and Using Internet*. Delhi: Global Business Press.

Nair, R. Raman (2002). *Accessing Information through Internet*. New Delhi: Ess Ess Publications.

Parekh, Harsha (1999). *Internet in the Scholarly Communication Process*. Mumbai: Mukherjee Knowledgeware Association.

Thulasi, K. and Rajashekar, T.B. (1999).Web Resources for Internet Use in Libraries. In: *Library and Information Networking: NACLIN 99* edited by H.K. Kaul. New Delhi: Delhi Library Network. Pp. 73-91.

Randall, Neil (2002). *Teach Yourself the Internet in a Week*. New Delhi: Prentice Hall of India.

Whittaker, Jason (2002) *Internet: Basics*. London: Routlage.

UNIT 15 INTERNET INFORMATION RESOURCES

Structure

- 15.0 Objectives
- 15.1 Introduction
- 15.2 Internet Information Resources
 - 15.2.1 Definition
 - 15.2.2 Nature of e-Resources
 - 15.2.3 Publishers of e-Resources
 - 15.2.4 Subscription of e-Resources
- 15.3 Types of Internet Resources
 - 15.3.1 General Information Resources
 - 15.3.2 Specific Information Resources
- 15.4 Searching the Internet : Where to Start
- 15.5 How to Keep Up-to-Date with New Internet Resources
- 15.6 Summary
- 15.7 Answers to Self Check Exercises
- 15.8 Keywords
- 15.9 References and Further Reading

15.0 OBJECTIVES

After reading this Unit, you will be able to:

- understand the definition and nature of Internet resources;
- know who the publishers of e-resources are and what subscription and access options are available;
- know various Internet resources in the area of Science and Technology and Medicine; Social Sciences; and Humanities;
- discover how to find information on the Net;
- find out how to critically analyse information resources;
- navigate the Internet resources successfully; and
- discover how to keep up-to-date with new websites / resources.

15.1 INTRODUCTION

You have studied in detail, in Block 4, Units 13 and 14 of this course, the vastness of the information resources available and the several applications and users of the World Wide Web (WWW). Utilising the web technologies that are becoming available the world over, various organisations are developing and maintaining websites providing access to the information generated/collected by them in the form of online databases, online publications, and also providing other information services and products.

This Unit will give details of the various information resources accessible on Internet and how to tap the right information resource in any desired rediscipline or topic of interest.

15.2 INTERNET INFORMATION RESOURCES

15.2.1 Definition

Internet information resources can be defined as resources that include both documents and non-documents in electronic or e-format that provide information or a pointer to the information and can be accessed via Internet.

15.2.2 Nature of e-Resources

These electronic resources could be of varied nature. Broadly, we could categorise them as follows:

- **Primary Sources of Information:** These include electronic conferences, electronic journals, electronic pre-prints and e-prints, electronic theses and dissertations, patents, standards, technical reports, project reports including status reports of current ongoing projects, news, software courseware, tutorials, manuals and the like.
- **Databases, Data sets and other Collections:** These include abstracting and indexing databases; digital collections comprising images, audio, video; scientific data sets comprising numeric, properties, structural databases; library catalogues; virtual libraries; museums and archives, etc.
- **Electronic Books:** Such as NetLibrary (<http://www.netlibrary.com/>); Ebrary (<http://www.ebrary.com/>), etc. Generally online book selling and print-on demand features also facilitated. For instance NetLibrary has entered into print-on-demand marketplace. Similarly Amazon.com (termed as the largest library — though not a library in true sense of the word) facilitates online book selling (<http://www.amazon.com/>)
- **Reference Sources:** such as dictionaries; encyclopaedias; biographies; handbooks; thesauri and the like.
- **Organisations and People:** Information about organisations and people ranging from funding agencies to libraries; information centres; research institutes; and experts; directories of people of varied nature (scientists; archaeologists, etc.)
- **Meta Resources:** Resources that facilitate easier access to network based resources in a defined subject area and a plethora of such resources under various names available on the Internet, such as subject gateways; virtual libraries; clearing house; pathfinders and the like.

15.2.3 Publishers of e-Resources

There is a mixed bunch of players in the market who publish electronic or e-information resources. These include traditional players offering electronic versions of their resources as well as several new enterprises offering new products and services that are 'born digital'. Besides, there are others also, such as subscription agents in their new role in the e-environment as electronic aggregators; institutions of higher learning and the like. We may broadly classify the Internet information resources publishers in the following categories:

- **Traditional Commercial Publishers:** Such as Elsevier Sciences' Science Direct (<http://www.sciencedirect.com/>); Springer Verlag (via their Link information services <http://link.springer.de>)

- **Scholarly Societies/Bodies:** Several such agencies are making electronic versions of their publications available online through their websites. For instance, Indian National Science Academy (INSA) scholarly publications are available online for free and unlimited access at <http://www.insa.ac.in> or via <http://www.insaindia.org>. Similarly, Indian Academy of Sciences, Bangalore journals can be accessed at www.ias.ac.in. Scholarly societies such as IEEE/IEE; ACM, etc., all are making e-versions of their publications available online.
- **Institutions of Higher Learning:** Several academic and research institutions host their specialised collections like courseware, theses and dissertations; and also offer subject gateways and portals, etc.
- **Electronic Aggregators:** These agencies facilitate intermediary services of aggregating electronic journals and other resources from different publishers and offer them to their clients through a single search system. Examples of aggregators are: JSTOR; OCLC; Lexis-Nexis and the like.
- **Others:** Among the others are non-commercial publishers, individuals, etc. Many individuals, groups and non-commercial publishers compile resources dedicated to providing access to e-resources (generally offered free) on the Internet. Examples of such meta resources include: Librarians' Index to the Internet (<http://lii.org/>); Vlib: The Virtual Library (<http://www.vlib.org/>); Internet Public Library (<http://www.ipl.org/>) and others.

15.2.4 Subscription of e-Resources

For production of a journal a significant investment on the part of publishers is involved irrespective of the format (print or electronic). Since publishers are adding value to the whole process of publication, somebody has to pay for it. So, journals are made available through the web at varying price models. Some of the common ones are as follows:

- **Electronic subscription bundled with print subscription:** In such cases electronic subscription to journals is linked with their print versions. This means, e-version may be either offered free with their print counterparts; or a fixed token percentage may be charged on and above the print subscription charges or some other similar model.
- **Electronic subscription with campus licenses:** wherein unlimited access to subscribed journals is facilitated campus-wide.
- **Electronic only:** In such situations publishers offer electronic versions only with a token discount.
- **Pay-per-look:** Identification of e-articles and paying for such articles in full text on a pay-per-look basis. This could be considered somewhat similar to print versions where articles of interest are identified and requested on payment basis via a document delivery system.
- **Consortium Licensing:** With the popularity of consortium mode of subscription, publishers have started working out best pricing models suitable for different types of consortia, whether at regional, local or national level.

Self Check Exercise

- 1) What do you understand by Internet information resources? Discuss the nature of electronic resources.
- 2) Who are publishers of e-resources on the Internet? Discuss giving suitable examples wherever possible.

- 3) Explain various ways of subscription of e-resources with special reference to journal subscription.

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit

.....

.....

.....

.....

.....

.....

15.3 TYPES OF INTERNET RESOURCES

The vast resources available on Internet have been broadly categorised type-wise for better understanding of them. However, since various services and resources provide for various facilities that are common, overlap of resources under various types is inevitable. It may be mentioned that only selected resources are given as examples for reference purposes, as it is neither feasible nor possible to include a complete list of them. The resources have been subjected to broad categorisation in order to make it comprehensible.

The resources can be grouped under two broad categories as follows:

- i) General, and
- ii) Specific

15.3.1 General Information Resource

In the changing electronic environment, libraries, which are the main facilitators of the scholarly communication, are caught in between the producers and the end-users. Though the researchers and information seekers usually tap the Internet resources from their desktops, yet many seek librarians' / information scientists' help for research information in the networked digital information environment. Users demand on librarians for more specific and relevant information is increasing in spite of librarian's and computer professionals' efforts for continuous improving the ease of access to information. To meet such increasing demands, library professionals have to keep on devising new strategies, methods and techniques for bringing together the information seekers and the information relevant to them on a continuing basis.

Under this category we shall be limiting to such resources which (a) are important for the library and information science professionals and (b) can facilitate you as navigators as to where to start. The following are some examples of this type:

- a) **Internet Library for Librarians** (<http://www.itcompany.com/inforetriever/>)

Internet Library for Librarians has been a popular information resource site for librarians since 1994. It is an information portal designed for librarians to locate Internet resources related to their profession. Internet Library for Librarians is also linked to and referenced by hundreds of libraries and other organisations, including the Library of Congress and OCLC. It provides links to over 3,000 resources. All the resources are recommended, selected, and reviewed by librarians. Each entry has a full description of the goals and /or scope of the resource, as well as the contact information, if provided. Internet Library for Librarians is a handy and useful tool for both novices and experienced library staff. It is a complete site in itself in the sense that it gives information about various types of resources at one place. There are resources ranging from library resources to associations, vendors, and the like. A snapshot of the homepage of the site is reproduced here which displays the types of resources the site facilitates access to.



Fig. 15.1: Screen Snapshot of Internet Library for Librarians Website

b) **Librarians' Index to the Internet (www.lii.org/)**

A superbly annotated and carefully selected and evaluated gathering of over 14,000 valuable resources on many topics. A view of the home page of the site is reproduced below for reference. The resources cover databases including invisible web searchable databases, directories/gateways, specific resources and the like.



Fig. 15.2: Screen Snapshot of Librarians Index to the Internet Website

c) **BUBL Information Service (http://bubl.ac.uk/)**

BUBL Link (http://bubl.ac.uk/link/)

It provides free user-friendly access to selected Internet resources covering all subject areas, with a special focus on library and information science. It has a large subject directory with British focus covering many academic and other topics.

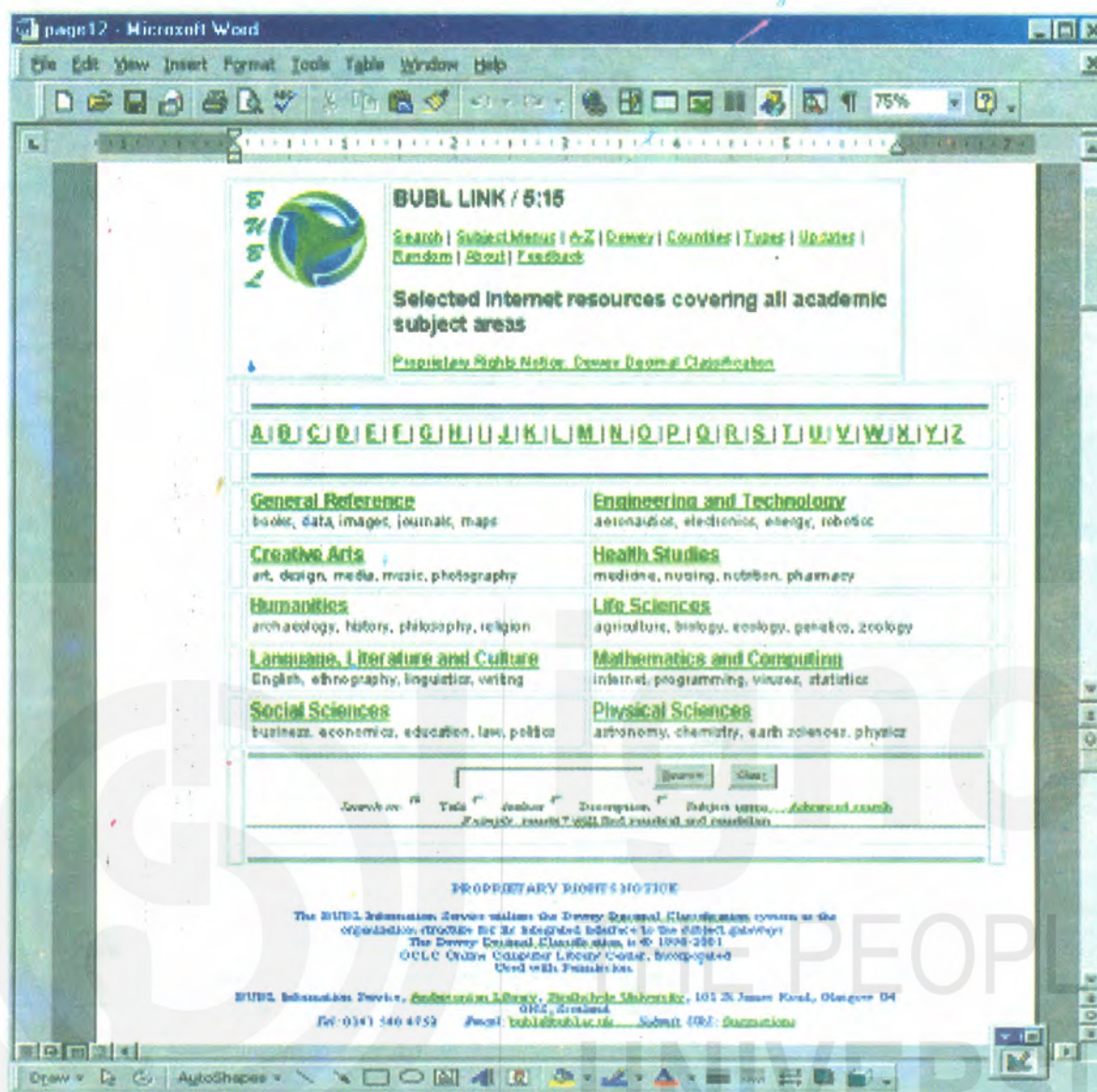


Fig. 15.3: Screen Snapshot of BUBL Link Website

d) **WWW Virtual Library (www.vlib.org)**

Virtual library facilitates access to many excellent resources. It is basically, a rich collection of subject guides on many subjects. However, sites are not annotated or evaluated.

e) **CIT-Online Resources for the Solo Librarian (http://www.unc.edu/cit/guides)**

It provides a collection of handy links for information professionals working in isolated or specialised libraries, or in other settings with limited access to a wide variety of on-site reference resources and other tools of the library profession. The site includes a great collection on: useful tools and services; book information and reviews; book sources (new, used and out-of-print); reference desks and virtual libraries; professional reading; professional organisations; librarian's marketplace, etc.

f) **Infomine (http://infomine.ucr.edu/)**

A collection of over 121,465 academically valuable resources, cooperatively compiled by university and college-level, academic librarians of the University of California campuses.

The following is the search page of the site:

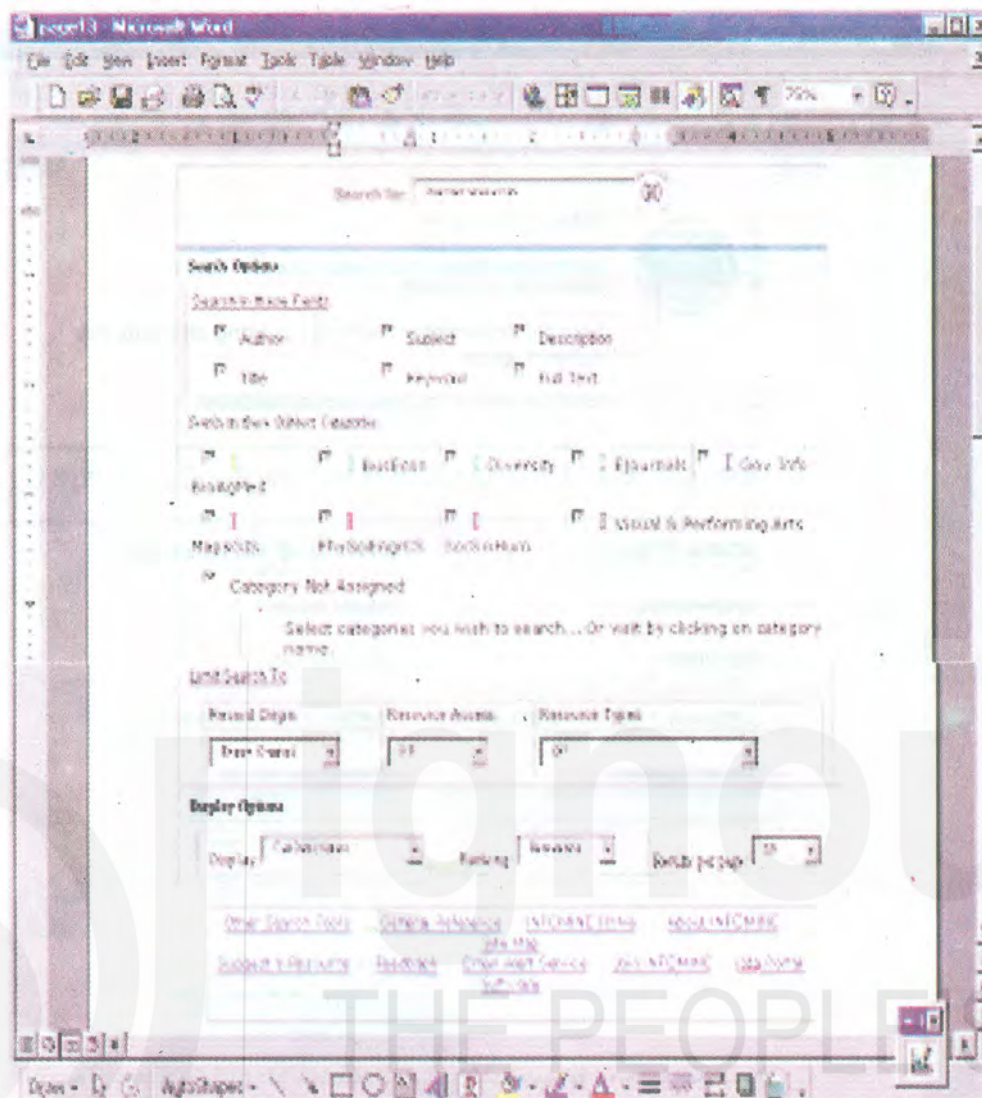


Fig. 15.4: Screen Snapshot of Search Page of Infomine

g) **Academic Info (www.academicinfo.net)**

Academic Info is a rich collection of web pages, databases, gateway pages, links to full-text publications, organisations selected primarily for academic research at the undergraduate level and above.

h) **The Internet Public Library Reference Center (www.ipl.org)**

A virtual reference library created in an attempt to replicate a library without walls on the Internet. Internet Public Library (IPL) is similar to using a reference room, with links to many resources by type and/or by subject.

i) **J-Gate (<http://j-gate.informindia.co.in>)**

This gateway to a new world of journals is a premier Indian portal for e-journals, conceived, developed and delivered by Informatics, India. It is a family of products and a kind of single source for librarians and information users to access share and manage their e-journals effectively. Among the high points of J-Gate are a directory of over 10,000 plus scholarly research and professional e-journals with links to journal and publisher sites; a Table of Contents (TOC) for several of these journals; and a searchable database, with links to full-text and to the Union Catalogue of journals in leading national libraries, to support resource sharing. Provides other facilities like - J-Gate Custom Content (JCC) - an exclusive local solution for management of the subscribed journals in the library of an individual institution or in a consortium mode; the other service being JCCC - a technology platform for sharing journal resources.

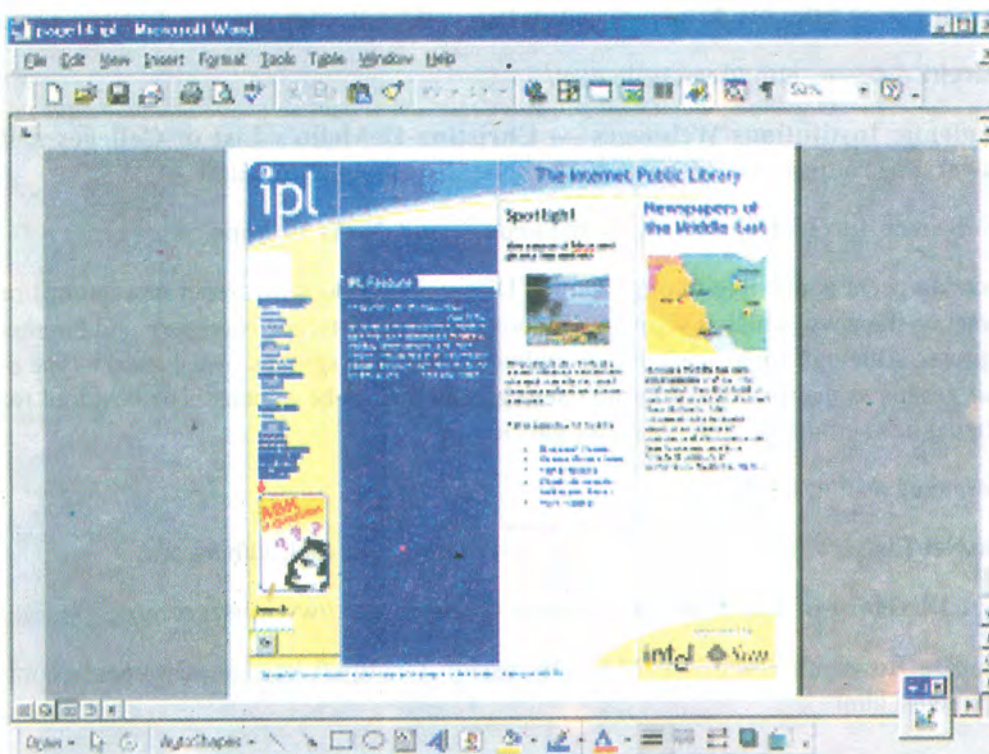


Fig. 15.5: Screen Snapshot of Internet Public Library Website

j) Other Sites

Other general sites for reference are given below:

i) General Search Sites

Google: <http://www.google.com/>

MetaCrawler: <http://www.metacrawler.com/>

HotBot: <http://www.hotbot.com/>

Alta Vista: <http://www.altavista.digital.com/>

Yahoo!: <http://www.yahoo.com/>

ii) People's E-mail Addresses/Web Pages

Search for Someone's E-mail Address, the following sources could be tried.

Yahoo! People Search: <http://people.yahoo.com/>

Internet @ddress Finder: <http://www.iaf.net/>

The White NetPages: <http://www.aldea.com/whitepages/white.html>

WhoWhere? E-mail Addresses: <http://www.whowhere.com/>

Bigfoot.com: <http://www.bigfoot.com/>

iii) Search for Someone's Telephone Number

The following site is useful to search for someone's telephone number.

Switchboard: <http://www.switchboard.com/>

AOL International Directories: <http://www.aol.com/netfind/international.html>

iv) College and University Web Pages

To search for College and University Web Pages, the following site may be tapped

SearchEdu.com: <http://searchedu.com/>

Academic Institutions Webpages — Christina DeMello's List of Colleges and Universities: <http://www.mit.edu:8001/people/cdemello/univ-full.html>

v) **Search for Listserv Lists, Web Forums, and News Groups**

Locating peers is possible through Online Discussion Groups or usenet newsgroups or usenet on the www which is supported by browsers line Netscape Navigator and Internet Explorer. Although some academic disciplines prefer Newsgroups over Listserv-type e-mail groups as their scholarly forum, Newsgroups tend to be popular. The best tool for locating newsgroups at present is 'Google Groups'.

For finding mailing lists

Tile.Net/Listserv: <http://www.tile.net/listserv/> is the best tool. Others are:

Liszt Directory of Email and Discussion Groups: <http://www.liszt.com/>

Impulse Research Corp.'s E-Mail Discussion Groups/Lists: <http://webcom.com/impulse/list.html>

Forum One's Forum Finder: <http://www.ForumOne.com/>

Stephanie da Silva's Publicly Accessible Mailing Lists: <http://paml.alastra.com/>

DejaNews Research Service: http://www.dejanews.com/home_ps.shtml

vi) **Maps**

Map Meta List: The Perry-Castañeda Library Map Collection: http://www.lib.utexas.edu/Libs/PCL/Map_collection/Map_collection.html

MapQuest: <http://www.mapquest.com/>

U. S. Census Map s: <http://tiger.census.gov/cgi-bin/mapbrowse-tbl>

Weather Maps: <http://www.provide.net/weather.html>

Maps of India: <http://www.natmo.org>

15.3.2 Specific Information Resources

Under the specific resources we shall be touching upon subject-specific resources under the three broad subject categories:

- a) **Science, Technology and Medicine (STM)**
- b) **Social Sciences**
- c) **Humanities**

Within these subject categories, we shall be discussing various information resources that have been further grouped under three broad groups so as to make them comprehensible. These groups are:

- By Source Type
- By Category
- By Organisation

Under each of these three groups, examples of various resource types for each of the subject category is included.

When the web first emerged, web sites consisted of relatively small collections of static, text-only data. The complexity of today's web sites reflects powerful new tools for developing and managing them. Scripting languages enable web developers to manage large amounts of data and create customisable interfaces that respond to user needs, while information architecture techniques ensure that web users quickly and efficiently locate information. With such developments taking place, the ease with which all types of resources get created and mounted on the Net, it is becoming unimaginable whether there are any types of sources that have been left out. In this Unit we shall take stock of the resources in Science, Technology and Medicine (STM); Social Sciences and Humanities by Source type.

There are several types of resources available such as:

- Electronic Journals and Newsletters
- Directories of all Types of WWW Electronic Journals
- Online Indexes of Print or Electronic Journals
- Table of Contents
- Preprints and Working Papers
- Discussion Lists or Forums/Usenet Newsgroups
- Directories of Newsgroups and Mailing Lists
- Software Archives
- Data Archives
- Subject Databases
- Campus Wide Information Systems (CWIS)
- Technical Reports
- Library Catalogues
- Patents
- Document Delivery
- Reference Sources
- Courseware Directories
- Other Resources

a) Electronic Journals and Newsletters

A journal, which is produced in an electronic format, is the electronic equivalent of a paper-based journal. Journals can be accessed over the Internet, which are appearing with increasing frequency and timeliness. Now an increasing number of journals are produced entirely in electronic format. Most of these journals are *full-text* journals. Such resources facilitate quick and easy access to current as well as back volumes. Besides, one does not need to go on browsing through all the articles, to pick up only papers of interest. There are number of sites that provide access to e-journal resources in Science Technology and Medicine (STM), Social Sciences and Humanities. Examples include:

i) Science Technology and Medicine (STM)

Elsevier Science (<http://www.elsevier.com/>) or (<http://www.sciencedirect.com>)

Elsevier Science has become the undisputed market leader in the publication and dissemination of literature covering the broad spectrum of scientific endeavors. It plays an important role in advancing the technologies necessary to create a seamless electronic

information delivery environment. The access to full-text data is fee-based. It provides access to over 1,800 scientific, technical and medical peer-reviewed journals; search to over 4 million articles and 59 million abstracts from scientific articles; and links to articles from over 120 other publishers, online access to multimedia features not otherwise available in print journals is also facilitated. In 2002 ScienceDirect launched a new commercial service for academic libraries, adding a new license for users to gain electronic access to both Elsevier Science and Academic Press journals on ScienceDirect. ScienceDirect E-Choice enables access via a single convenient platform, and a single license agreement. Flexible access for ScienceDirect guest users - that is, users not associated with a ScienceDirect account - are permitted to browse and read abstracts from all of the ScienceDirect journals for free. They can also set up free table of contents e-mail alerts and create personal journals.

Elsevier Science is a world leading publisher of scientific, technical and medical information

- FOR AUTHORS
 - Author Gateway
- FOR EDITORS
 - Information and Services for Editors
- FOR READERS
 - Journal Table of Contents
 - Contents Direct
 - Books Butler
 - Newsletters and Other Services
- FOR LIBRARIANS
 - E-Products
 - 2002 Subscription Price List
 - Journal History/Title Changes

Search our products

ELSEVIER SCIENCE PREMIER ELECTRONIC PRODUCTS

The premier source for STM information

What makes doctors click

An overview of Elsevier Science electronic products

Search our full text articles

powered by **SCIFUS** Search Engine for Science

"Minutes of science" kept in perpetuity as Elsevier Science and Koninklijke Bibliotheek finalise major archiving agreement

In an historic and groundbreaking initiative, Elsevier Science and the Koninklijke Bibliotheek (National Library of the Netherlands) have finalised an agreement, establishing the Library as the first official digital archive for Elsevier Science journals. Full story, with background and photos.

Medical reference site for French-speaking doctors
A new medical reference site, in

Fig. 15.6: Screen Snapshot of Elsevier Science Website

2002 Books Price List

New and Forthcoming
Publications

FOR AGENTS/
BOOKSELLERS

Published books

Dispatch dates

2002 Subscription Price List

2002 Books Price List

New and Forthcoming
Publications

FOR PRESS

Media Newsletter

Company News - archive

CAREERS

Getting to know us

Working with us

Job Search

French, was recently launched by Editions Scientifiques et Medicales. Entitled "emc-consulte.com", the site provides subscribers with access to medical information taken from Encyclopédie Médico-chirurgicale, Elsevier Science's medical encyclopedia for the Latin language market. Visit site.

ScienceDirect supports Athens authentication

ScienceDirect® users from universities and colleges throughout the UK can now benefit from remote off-campus access to ScienceDirect (a division of Elsevier Science) without additional administration by those institutions. ScienceDirect now supports the Athens® Access Management System in the UK. More details.

Fig. 15.6: (contd.) Screen Snapshot of Elsevier Science Website

Springer Science Online (<http://www.springer.de/>)

Springer publishes annually over 4,000 new books and approximately 700 journals, most of which are available in electronic form.

Blackwell Scientific Journals (<http://www.blacksci.co.uk/uk/journals.htm>)

Publishes over 600 prestigious journals. Nearly all Blackwell Publishing journals are available online as well as in the print edition.

ii) Social Sciences

World Wide Web Virtual Library - Social Sciences (<http://vlib.org/SocialScience.html>)

It includes over 100 journals on various areas of Social Sciences. The index lists journals in alphabetical order and search can be initiated by clicking on the letter to jump to the appropriate section. Besides, directories of all types of WWW e-Journals and other online indexes of print or electronic journals are also part of this site. A snapshot of the home page World Wide Web Virtual Library showing various subject categories for which resources are available is reproduced below. On clicking on the concerned link it shows a pool of resources in the area. Sites of journals included in the Social Sciences World Wide Web Virtual Library as such cannot be reproduced, since it is a long list, which you will be able to explore by yourself on navigating this site.

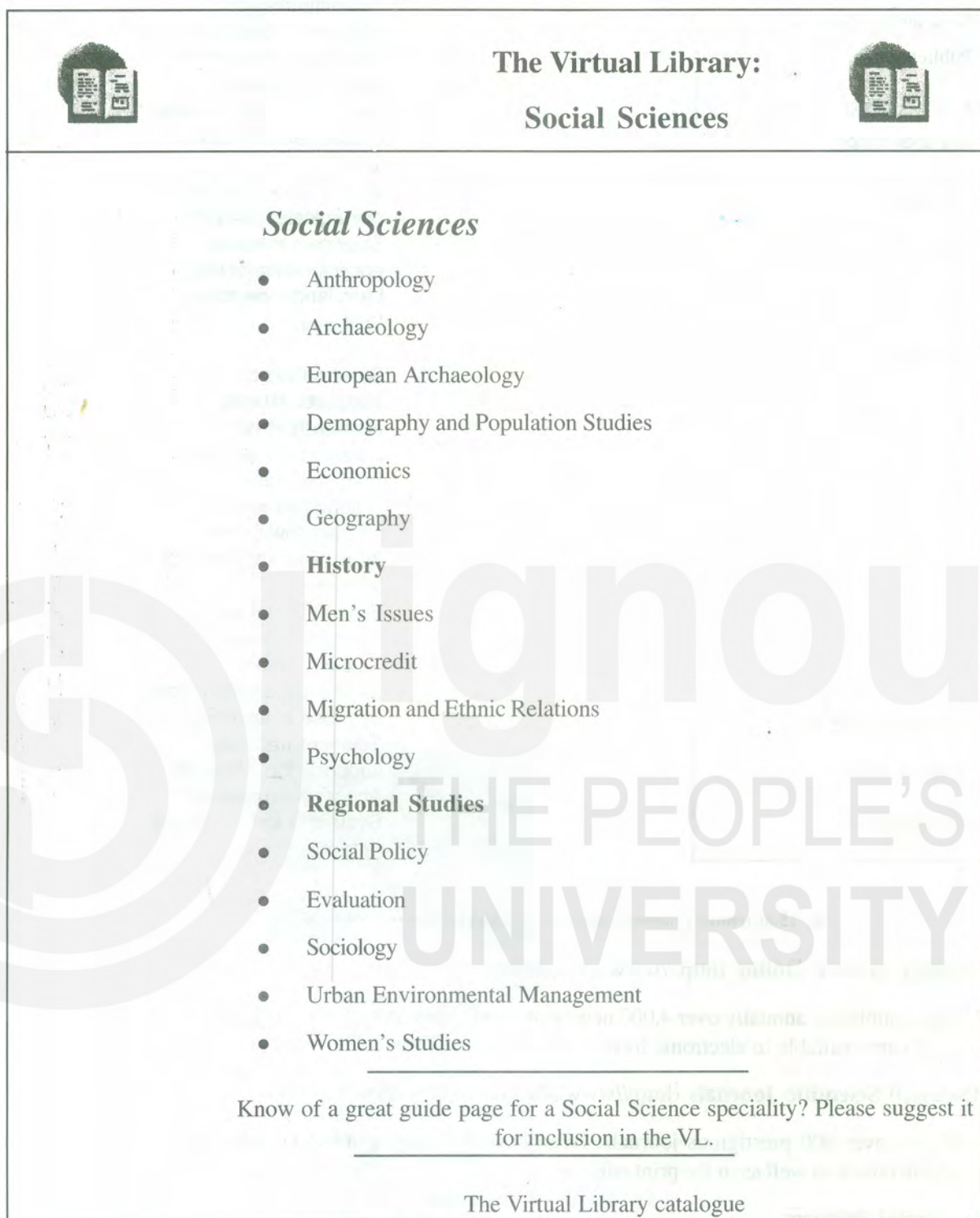


Fig. 15.7: Screen Snapshot of The Virtual Library Website

Register of Leading Social sciences E-Journals

(<http://www.clas.ufl.edu/users/gthursby/socsci/ejournal.html>)

Keeps track of on-line serials of significance to researchers in Social Sciences and Humanities. This site is part of the Social Sciences Virtual Library, which was established at the Australian National University.

ECONbase- Elsevier Science

(<http://www.elsevier.com/homepage/sae/econworld/menu.htm>)

Or (<http://www.elsevier.nl/>)

This site acts as an access point to 79 economics journals. Currently, the resource-base provides access to 60,000+ online papers. Online access to full-text articles in ECONbase is available to those readers whose library is - either registered with ScienceDirect web editions or subscribes to ScienceDirect Digital Collections. In all cases, access is restricted to those journals to which the library holds a current subscription. Except for online access to full-text, there are no restrictions on access to information within ECONbase. Available functionality includes: browsing tables of content, searching titles, authors, abstracts, keywords in a database of 67 journals. Besides this, provision is also made available for author and keyword indexes for each individual journal, viewing abstracts for all 67 ECONbase journals, selected full-text of a limited number of journals and a free sample copy of each journal.

PSYCLINE (<http://www.psycline.org/journals/psycline.html>)

The website was started in 1995 by Dr. Armin Günther a psychologist, from Germany, under its former name **Links to Psychological Journals** and has won a high reputation as one of the most comprehensive and up-to-date index of psychology and social science journals on the web. It provides access to over 1,500 journals in the area of Psychology and Social Sciences with free table of contents and abstracts.

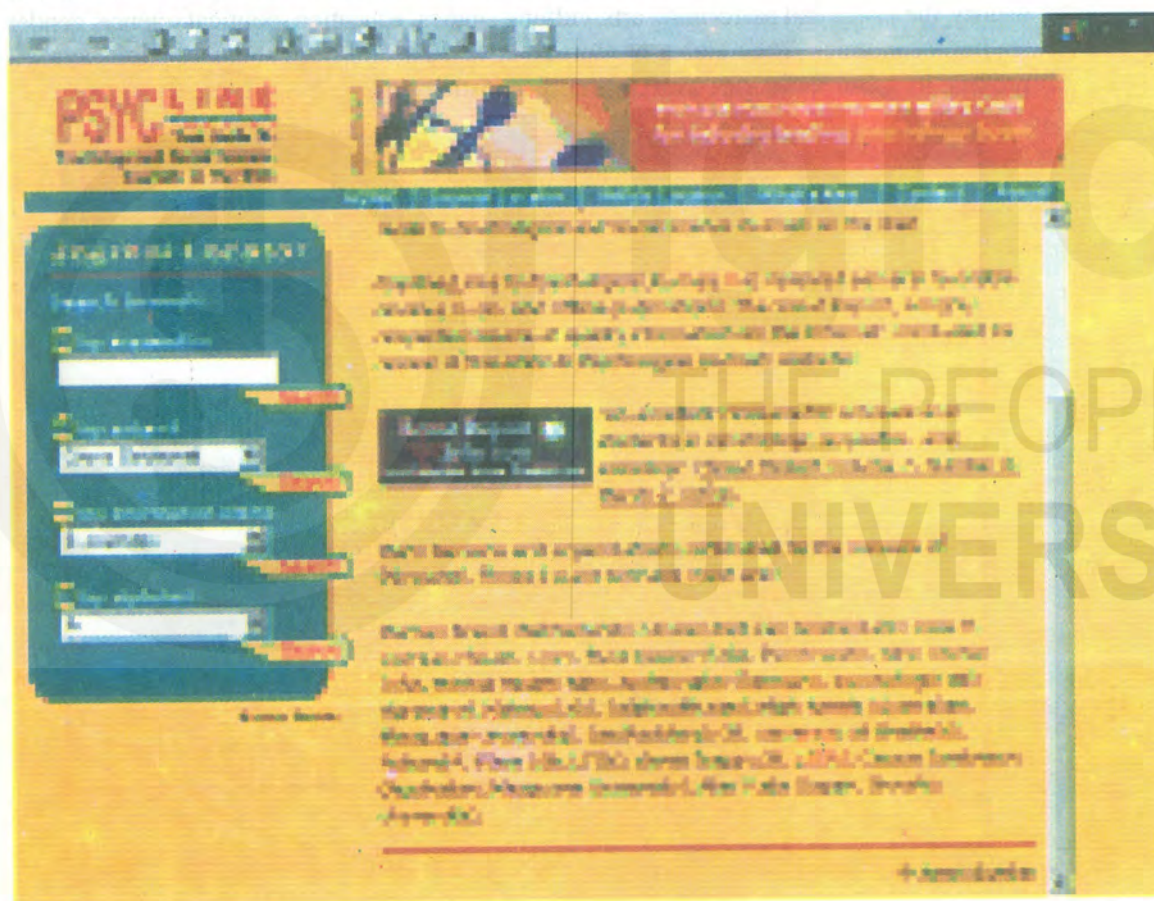


Fig. 15.8: Screen Snapshot of PSYCLINE Website

iii) Humanities

Technical Documentation from HW Wilson - Humanities Index/Abstracts with Full-Text Journal List (<http://www.hwwilson.com/Databases/humani.htm>)

It is a good source for tracking journals in the Humanities area. Humanities Index/Abstracts with Full Text over 550 journals, excluding name changes. This number represents active, ceased, dropped journal titles.

Social Sciences and Humanities Electronic Journal -UCSD Libraries (http://libraries.ucsd.edu/sage/ejournals/social_sciences_and_humanities.html)

It is another source that provides access to a large number of e-journals in the area of Humanities and Arts.

CARL UNCOVER (<http://www.ingenta.com/>)

CARL is a computerised network of library services developed by the Colorado Alliance of Research Libraries (CARL). CARL UnCover is the Alliance's index to journals and magazines. Access to CARL UnCover is now available through the web at Ingenta. Ingenta has been working to integrate the two databases in order to provide you with a more comprehensive and easy-to-use service. The integrated service offers free searching and browsing of more than 27,000 publications with 11,000 titles that were not available in UnCover. In addition, a number of new services have been introduced. Searching Ingenta is free but Article delivery service is charged. One can arrange to have articles faxed, delivered, or sent electronically (HTML or PDF format) directly. One has to provide his personal credit card number and the articles will be charged to that account.

Project Muse (<http://muse.jhu.edu/>)

Project MUSE is a joint venture of Johns Hopkins University Press and Milton S. Eisenhower Library. It offers access to journal titles in the field of literature and criticism, history, the visual and performing arts, cultural studies, education, political science, gender studies, economics. The number of journals covered is 250 from about 40 publishers. At present, its subscriptions are available only to institutions.

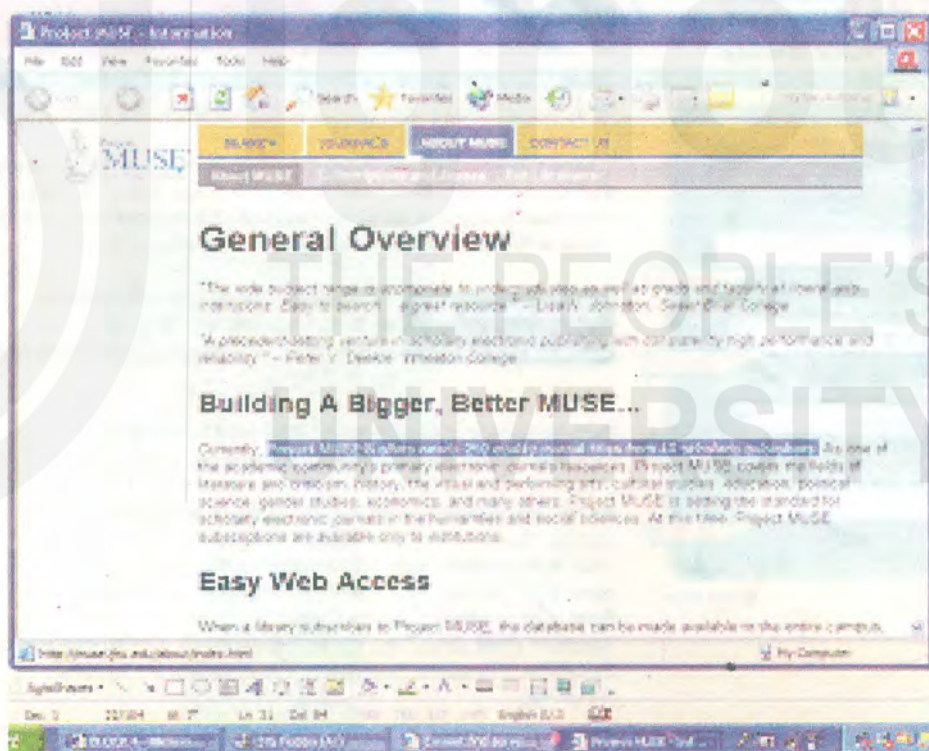


Fig. 15.9: Screen Snapshot of Project MUSE Website

ISI Web of Science (<http://wos.mimas.ac.uk/>)

ISI Web of Science service is a massive resource base covering all the three main divisions of human knowledge, i.e., Science and Technology, Social Sciences and Humanities. The basic databases covered under Web of Science and made available are:

- Science Citation Index Expanded with Cited References and Author Abstracts (1981-)
- Social Sciences Citation Index Expanded with Cited References and Author Abstracts (1981-)
- Arts and Humanities Citation Index with Cited References (1981-)

b) **Directories of All Types of WWW E-Journals**
World Wide Web Virtual Library (<http://www.vlib.org/>)

It is the oldest catalogue of resources on the web. It was started by none other than the creator of WWW, Tim Berners-Lee in 1991. It covers all the branches of knowledge ranging from arts to social sciences, sciences, engineering, and technology. It is an excellent source for starting the search for e- journals.

Directories of Electronic Journals (<http://gort.ucsd.edu/ejournal/jdir.html>)

This site provides a keyword searchable database and extensive links to other resource facilities, supported by the University of California at San Diego.

New Jour (<http://gort.ucsd.edu/newjour/>)

It is an archive for a major list of electronic journals and newsletters available on the Internet.

Scholarly Journals Distributed via the World Wide Web (U. Houston, USA)

(<http://info.lib.uh.edu/wj/webjour.html>)

It is a directory organised alphabetically that provides links to established web-based scholarly journals that offer access to English language article files for free.

c) **Online Indexes of Print or Electronic Journals**

Social Sciences World Wide Web Virtual Library also includes other online indexes of print or electronic journals for reference purposes. Examples include the following:

Anthropological Index Online

(http://lucy.ukc.ac.uk/cgi-bin/uncgi/Search_AI/search_bib_ai/anthind)

An index to current periodicals in the Museum of Mankind Library (incorporating the former Royal Anthropological Institute library). Currently it is available for the years from 1970 onwards. The index is searchable by year, subject area, author, title, or journal.

d) **Table of Contents (TOC)**

In order to keep abreast of new developments in a particular topic or issue, one of the means or services that people use is current contents listings of the journals. With the advent of e-publishing and Internet, these current contents of journals have become current in real time situations. Not many journals have full-text available on the web. Hence, many academic publishers have designed web pages for each of their journals, wherein they list the current contents of the latest issues of the journal, sometimes with abstracts. So, a regular scan of relevant journal web pages can help to keep up-to-date with new articles as they appear. Almost all the journals available in e-format provide access to table of contents for free.

This is a very useful resource for librarians for Current Awareness Service. These resources help users to stay up-to-date in their research. Also, all such resources are an excellent aid in saving research time. Examples include:

i) **STM**

- 'Contents Direct' Service by Elsevier (<http://www.contentdirect.elsevier.com/>): the service covers over 800 journals. 'Uncover' service provided by CARL agency having TOC of over 16,000 journals. ISI's TOC Alerting Service (journal tracker) (http://alerting.isinet.com/jt_home.html). It provides a choice of about 8,000 journals from which you are allowed to choose 25 titles. You are provided with the complete bibliographical details besides the author abstracts.

ii) **Social Sciences**

Academic Press Journals (<http://www.apnet.com/journals>)

Now part of Elsevier Science provides links to the latest Table of Contents, and subscription information for all 174 Academic Press journals in the area of social and behavioural sciences.

iii) **Humanities**

Current Contents / Arts and Humanities

(<http://www.isinet.com/isi/products/cc/editions/ccah/>)

ISI Current Contents/Arts and Humanities provides access to complete bibliographic information from articles, editorials, meeting abstracts, commentaries, and all other significant items in recently published editions of over 1,120 of the world's leading arts and humanities journals and books from a broad range of categories. The source besides facilitating regular features of such service has provision for combining comprehensive coverage with numerous access points, exclusive search capabilities; optional coverage of past research; and saves research time by providing one source for a variety of research information - including author abstracts, author addresses, and more information per bibliographic record than in other resources.

Current Contents Search

(<http://library.dialog.com/bluesheets/html/bl0440.html>)

A weekly service that reproduces the tables of contents from current issues of leading journals in the arts and humanities, social sciences, and sciences.

e) **Preprints and Working Papers**

The term 'preprint' most often refers to a manuscript that has not yet been published, but may have been reviewed and accepted, submitted for publication, or intended for publication and being circulated for comments. A preprint accessible over the web may also be referred to as an 'e-print'. Many e-prints are electronic versions of research papers that have been submitted for dissemination and review among peers; for publication in journals; or prior to presentation at conferences. Preprints also cover papers that authors have submitted for journal publication, but for which no publication decision has been reached, or even papers electronically posted for peer consideration and comment before submission for publication. In fact, preprints can also be documents that have not been submitted to any journal. Some preprint servers may define preprints as any electronic work circulated by the author outside of the traditional publishing environment.

The Internet is increasingly being used by academics to publish the full-text of conference papers, draft papers or work-in-progress, and other similar material, often to facilitate peer review process. Examples include the following:

i) **STM**

ePrint Network (<http://www.osti.gov/eprints/index.html>)

Searchable gateway to eprint servers that deal with scientific and technical disciplines of concern to DOE. The ePRINT Network provides access to electronic preprints available from diverse sites. Developed by the U.S. Department of Energy (DOE) Office of Scientific and Technical Information (OSTI), the Network is a "one-stop shopping" site for eprints in science and technology.

Directory of Mathematics Preprint and e-Print Server (American Mathematical Society) (<http://www.ams.org/global-preprints>)



Fig. 15.10: Screen Snapshot of Directory of Mathematics Preprint and e-print Servers Website

PubMed Central (<http://www.pubmedcentral.nih.gov/>)

PubMed Central is a digital archive of life sciences journal literature, developed and managed by the National Center for Biotechnology Information (NCBI) at the U.S. National Library of Medicine (NLM). It contains data from diverse sources stored in a common format in a single repository. It allows fast searching, manipulation and cross-linking of the complete collection. PubMed Central provides the following two services: GenBank, the genetic sequence data repository, and PubMed, the database of citations and abstracts to biomedical and other life science journal literature. GenBank, along with the powerful search facilities has helped molecular biologists make rapid advances in their field. PubMed (which encompasses Medline) helps, to locate relevant articles and, in many cases, link directly to a publisher's site for the full-text. GenBank has proven the advantages of collecting DNA sequences in a central repository with a common format.

ii) Social Sciences

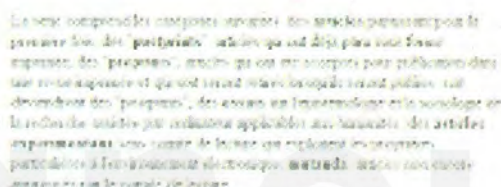
WoPEc (<http://netec.mcc.ac.uk/WoPEc.html>)

It is an effort at international level to collect together and make available working papers in the area of economics from academics. WoPEc is part of a larger project called NetEC, an international academic project for networking interactions in economics. It provides access to thousands of working papers from hundreds of series and contains over 100,000 documents in electronic format: 76,311 working papers; 100,286 journal articles and 3,900 registered authors. These are downloadable but not necessarily free.

iii) Humanities

CH Working Papers <http://www.chass.utoronto.ca/epc/chwp/>

CH Working Papers (or Computing in the Humanities Working Papers) provides access to refereed articles of an interdisciplinary nature on computer-assisted research. They serve the needs of the researcher at an intermediary stage at which questions of computer methodology in relation to the main discipline are of interest to the scholar before the computer disappears into the background. A screen shot of the site is reproduced here:



456

Examples include the following:

i) **STM**

Gentalk - Subscription to: listserv@usa.net

Provides a forum for discussion of genetic problems, lab protocols, current issues dealing with genetics and genetic engineering in general.

ii) **Social Sciences**

Mailbase (<http://www.mailbase.ac.uk/lists.html>)

Provides access to over 2,000 electronic discussion lists for the UK higher education and research community. The site provides education descriptions, message archives and subscription information for numerous mailing lists relevant to education. The site also includes information about how to join any of these lists.

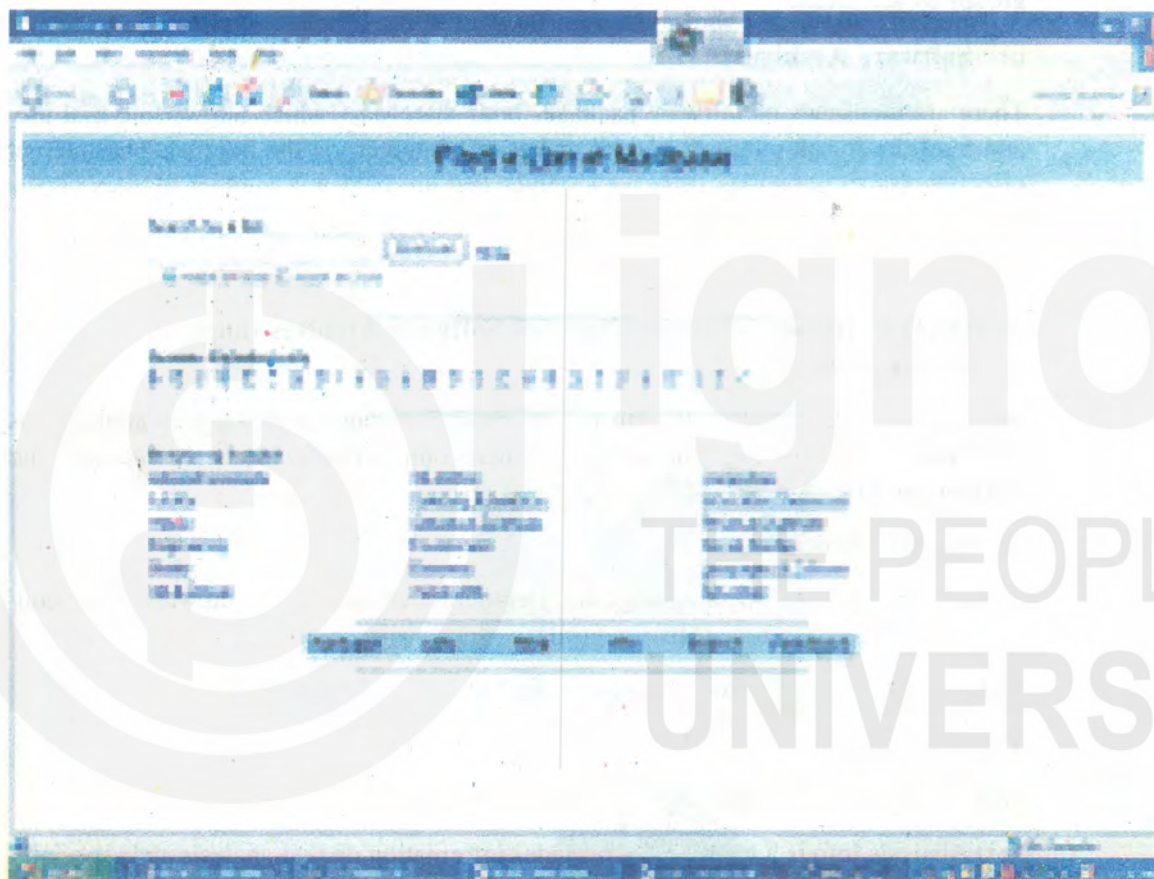


Fig. 15.12: Screen Snapshot of Mailbase Website

The SOSIG Mailing List (<http://www.mailbase.ac.uk/lists/sosig/>)

It has over 400 members from the worldwide social science community and distributes messages about among other things, new Internet sites and services for social scientists.

iii) **Humanities**

HUMBUL-Humanities Bulletin Board (<http://users.ox.ac.uk/~humbul>)

Gateway site, maintained by Chris Stephens at Oxford University, is the best Internet resource in the humanities, and also with a conference diary. Search or browse by clicking on 19 broad categories in the arts and humanities. A good starting point to identify quality web resources in this area.

Humanist Discussion Group (<http://www.princeton.edu/~mccarty/humanist/humanist.html>)

Exclusively for people working and interested in Humanities related discussions.

g) Directories of Newsgroups and Mailing Lists

Deja.com (formerly Deja News) (<http://www.deja.com>)

Provides access to a vast number of discussion forums and Usenet groups, including archives of previous postings. The new Deja.com now also aims to serve as an Internet consumer guide, with ratings of products. Deja Tracker informs you by e-mail about new postings in your favourite newsgroups.

ForumOne (<http://www.forumone.com>) Speciality search engine; helps you locate messages posted on over 270,000 web discussion forums.

Liszt (<http://www.liszt.com>) Searchable database of over 90,000 mailing lists/discussion groups.

Usenet Groups (<ftp://rtfm.mit.edu/pub/usenet-by-hierarchy/>) A directory to Usenet groups by hierarchy.

h) Software Archives

There are thousands of software packages both shareware, which allows free trial use, and freeware for all purposes and all makes of computer via the Internet. The Internet offers access to numerous software archives that are held at a number of sites on the Internet. Examples include the following:

i) STM

HENSA-The Higher Education National Software Archives (<http://www.hensa.ac.uk/>)

Software is mostly in public domain and shareware covering a wide range of applications but especially networking. The archive of microcomputer software is at Lancaster, and the Unix archive at Kent.

ii) Social Sciences

Software and Datasets for Sociology and Demography (http://www.stat.washington.edu/raftery/Research/Soc/soc_software.html)

There are several software and datasets available from this site.

iii) Humanities

Text Analysis Info Page (<http://www.textanalysis.info/>)

Text Analysis Info is a website that provides information on text analysis and especially software for the analysis of human communication content. This is mostly text, but not limited and quite a few programs can handle audio and/or video data. The first version started in April 1999, has become quite popular site now. There is a whole lot of information and there are links to various resources available at this site ranging from conferences, workshops and forums; mailing lists; news to text archives; books and regressive imagery dictionary and of course, to software (classification of text; analysis software; definitions and terms; transcribing software (audio/video); language: linguistics information retrieval and much more).

i) Data Archives

A data archive is a permanent, electronic collection of datasets with accompanying metadata such that users of the data can acquire, understand, and use the data. Data archives are resource centres for analysts who use data for research and teaching. Data archiving is a method of conserving very expensive resources and ensuring that their research potential is fully exploited. Archives ensure that when technology changes, the data in their holdings are technically transformed to remain readable in the new environment. Their functions usually include: being more than a long-term backup, being more than an

index or catalogue with pointers to datasets stored elsewhere, ensuring that data are preserved against technological obsolescence and physical damage, cataloguing their technical and substantive properties for information and retrieval, supplying them in an appropriate form to secondary users. Data archives have been established in most European countries and in the United States. They are actively used for testing hypotheses and for other scholarly purposes. Examples include the following:

i) **STM**

National Space Science Data Center (<http://nssdc.gsfc.nasa.gov/>)

The National Space Science Data Center (NSSDC), USA archives and provides access to a wide variety of astrophysics, space physics, solar physics, lunar and planetary data from NASA space flight missions, in addition to selected other data and some models and software. NSSDC provides online information bases about NASA and non-NASA data as well as the spacecraft and experiments that have or will provide public access data. NSSDC also provides information and support related to data management standards and technologies and much more.

ii) **Social Sciences**

There is large number of data archives available on the Internet in the area of social sciences. Some of these are indicated below:

Social Science Hub (<http://www.sshub.com/index.html>)

Social Science Hub covers resources for Anthropology, Sociology and Archaeology and other associated disciplines. Provides links to data archives, websites, newsgroups, news, research tools and publications

Guide to Primary Social Science Research Data and Related Resources Available on the Internet (<http://www.chass.utoronto.ca/datalib/other/>)

It is a guide to data libraries, data archives, and related institutions about which information is available through the Internet, as well as to primary research data and related resources available for access or acquisition. Literature is also available on data management. The guide includes quantitative (numeric) data, as well as textual resources. It is a major resource guide to provide links to directories of data archives and data libraries; individual data archives, data libraries, and related institutions, besides other useful resources in social sciences. Broadly it covers the following:

- 1) Directories of data archives and data libraries
- 2) Indices, union catalogues, and resource guides to research data files
- 3) Professional data associations
- 4) Individual data archives, data libraries, and related institutions
- 5) Data producers
- 6) Other data resources
- 7) Statistics and software, documentation and sources
- 8) Text archives
- 9) Selected text resources
- 10) Data-related conferences
- 11) Data-related training opportunities
- 12) Electronic journals

13) Electronic listservs.

Clickable Map of Major Social Science Data Archives

(<http://www.nsd.uib.no/cessda/europe.html>)

Provides links to various social science data archives of Europe.

The Data and Program Library Service (DPLS) (<http://dpls.dacc.wisc.edu/>)

It is the central repository of data collections used by the social science research community at the University of Wisconsin-Madison. Its mission is to promote academic research by facilitating the use of secondary research materials. To fulfill this mission the library acquires, preserves and facilitates access to social science data resources, provides reference and technical services to researchers, and assists in the archiving of locally produced data. The DPLS is part of the Data and Computation Center (DACC) and is subject to its policies. A faculty advisory committee works with the Director of DACC, and DPLS staff, to oversee policy.

The data library's current holdings mainly support social science research employing statistical methods on numeric data. These holdings include several thousand studies. The majority of these studies were obtained through the University of Wisconsin-Madison membership in the Inter-university Consortium of Political and Social Research (ICPSR). They also acquire data from inter-governmental organisations, U.S. statistical agencies, and other data sources. Their collection spans a wide range of topics including historical and contemporary population characteristics, community and urban studies, intra- and international conflict, economic behaviour and attitudes, education, mass political behaviour and attitudes, and social institutions and behaviors. The data library facilitates access to its data collection via computer-based information resources such as the World Wide Web, through teaching classes both in-house and in the classroom, and in one-on-one situations. The goal is to provide access to data as quickly and efficiently as possible while at the same time meeting the individual needs of the user. Whenever possible and relevant, users are provided with electronic versions of study documentation and input syntax for SPSS or SAS.

Social Science Data Archives (Australia) (<http://assda.anu.edu.au/>)



The Social Science Data Archives (SSDA), was set up in 1981 at the Research School of Social Sciences at the Australian National University. It provides access to computer-readable data relating to social, political and economic affairs of Australia. Its Home Page is shown in Fig. 15.13.

iii) Humanities

JSTOR (www.jstor.org)

Provides full-texts of articles relating to ecology, economics, education, finance, history, mathematic, political science, and population studies. Resources are listed by subject and by title also.

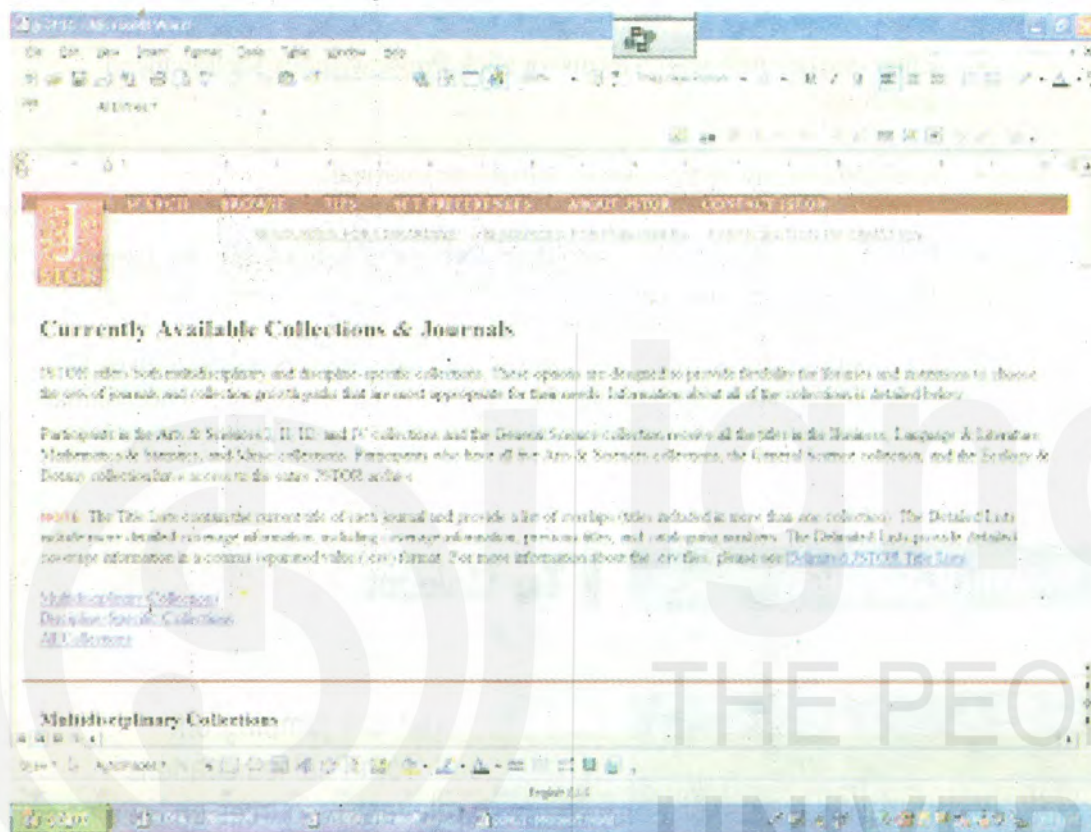


Fig. 15.14: Screen Snapshot of JSTOR Website

Arts and Humanities Data Services (<http://www.ahds.ac.uk/>)

The Arts and Humanities Data Services (AHDS) is a UK national service funded by the Joint Information Systems Committee and the Arts and Humanities Research Board. AHDS helps you to create, deposit, preserve or discover and use digital collections in the arts and humanities. It provides access to the work of the AHDS, their training events, publications, and wide-ranging collections.

j) Subject Databases

Databases are a collection of records each of which contains details of a different data item, whether numeric, textual or image-based, and which are usually available in a searchable format. Examples of database resources available on Internet include the following:

i) STM

PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>)

PubMed is one of the several versions of MEDLINE made available via the Internet. The database provides access to nine million citations held in original MEDLINE database plus pre-MEDLINE basic data which has not yet been added to MEDLINE). Besides, the site also provides access to other related databases.



- Enter one or more search terms, or click [Preview/Index](#) for advanced searching.
- Enter author names as smith jc. Initials are optional.
- Enter journal titles in full or as MEDLINE abbreviations. Use the [Journal Browser](#) to find journal titles.

PubMed, a service of the National Library of Medicine, provides access to over 12 million MEDLINE citations back to the mid-1960's and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources.

Bookshelf Additions

Developmental Biology, 6th ed.

by SF Gilbert

Surgical Treatments - Evidence Based and Problem-Oriented

edited by RA Holzheimer & JA Mannick

and NCBI's Genes and Disease.

These three resources are now available for interactive searches on [Bookshelf](#).

Try LinkOut



[LinkOut](#) is a jumping-off point from PubMed citations to relevant resources on the web, such as, full-text articles, library holdings, commentaries, author biographies, practice guidelines, consumer health information, and research tools. [Example LinkOut resources](#) are available.

[Write to the Help Desk](#)

[NCBI](#) | [NLM](#) | [NIH](#)

[Department of Health & Human Services](#)

[Freedom of Information Act](#) | [Disclaimer](#)

UNESCO Social Science Database - DARE: Directory of Social Sciences Institutions, Specialists, Periodicals (<http://www.unesco.org/most/dare.htm>)

The DARE database offers over 11,000 worldwide references to social science research and training institutes; social sciences specialists; social science documentation and information services; social science periodicals. The database also contains special references to peace, human rights and international law research institutes.

ERIC- Educational Resources Information Center (<http://www.eric.ed.gov/>)

The ERIC database is the world's largest source of education information. The database contains more than one million abstracts of education-related documents and journal articles. You can access the ERIC database on the Internet or through commercial vendors and public networks. You can also access ERIC abstracts in the print publications Resources in Education and Current Index to Journals in Education.

By searching ERIC database, you will retrieve citations and abstracts for education-related literature relevant to your search topic. These citations and abstracts are called 'resumes'. There are two types of resumes in the ERIC database: ERIC Documents (ED) and ERIC Journal Articles (EJ). The homepage of ERIC is reproduced below:

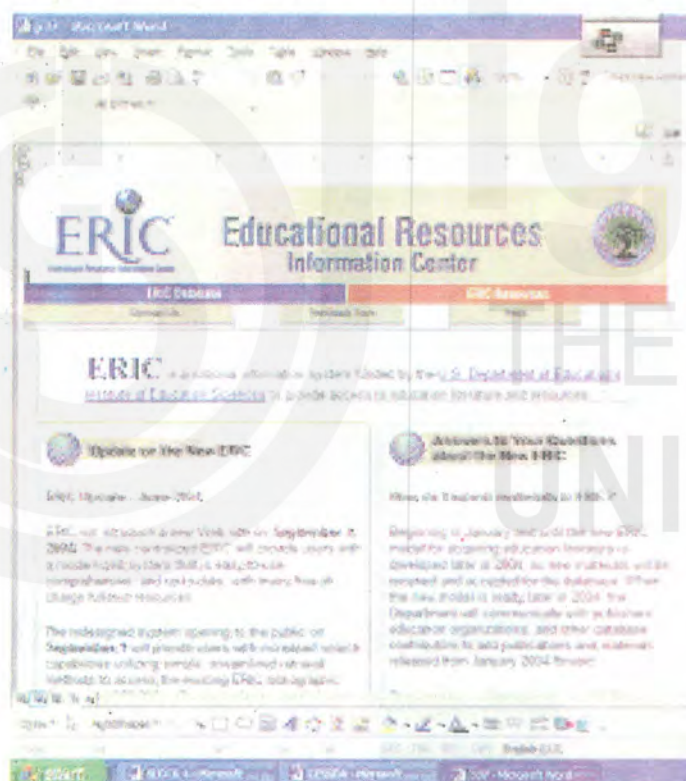


Fig. 15.16: Screen Snapshot of ERIC Website

Population Index on the Web (<http://popindex.princeton.edu/>)

Population Index is the primary reference tool to the world's population literature. It presents an annotated bibliography of recently published books, journal articles, working papers, and other materials on population topics. This website provides a searchable and browsable database containing 46,035 abstracts of demographic literature published in Population Index in the period 1986-2000.

The SSRN (Social Science Research Network) (<http://www.ssrn.com/>)

Social Science Research Network (SSRN) is devoted to the rapid worldwide dissemination of social science research and is composed of a number of specialised research networks in each of the social sciences. Each of SSRN's networks encourages the early distribution of research results by publishing submitted abstracts and by soliciting abstracts of top

quality research papers around the world. There are now hundreds of journals, publishers, and institutions - academic and other cooperating institutions - that provide working papers for distribution through SSRN's eLibrary and abstracts for publication in SSRN's electronic journals. The SSRN eLibrary consists of two parts: an abstract database containing abstracts of over 42,600 scholarly working papers and forthcoming papers and an electronic paper collection currently containing over 23,600 downloadable full-text documents in Adobe Acrobat PDF format. The eLibrary also includes the research papers of a number of fee-based publications. The network encourages readers to communicate directly with authors and other subscribers concerning their own and others' research. SSRN's database abstracts of some journals for both accepted paper series and working paper series. The homepage of SSRN is reproduced below:

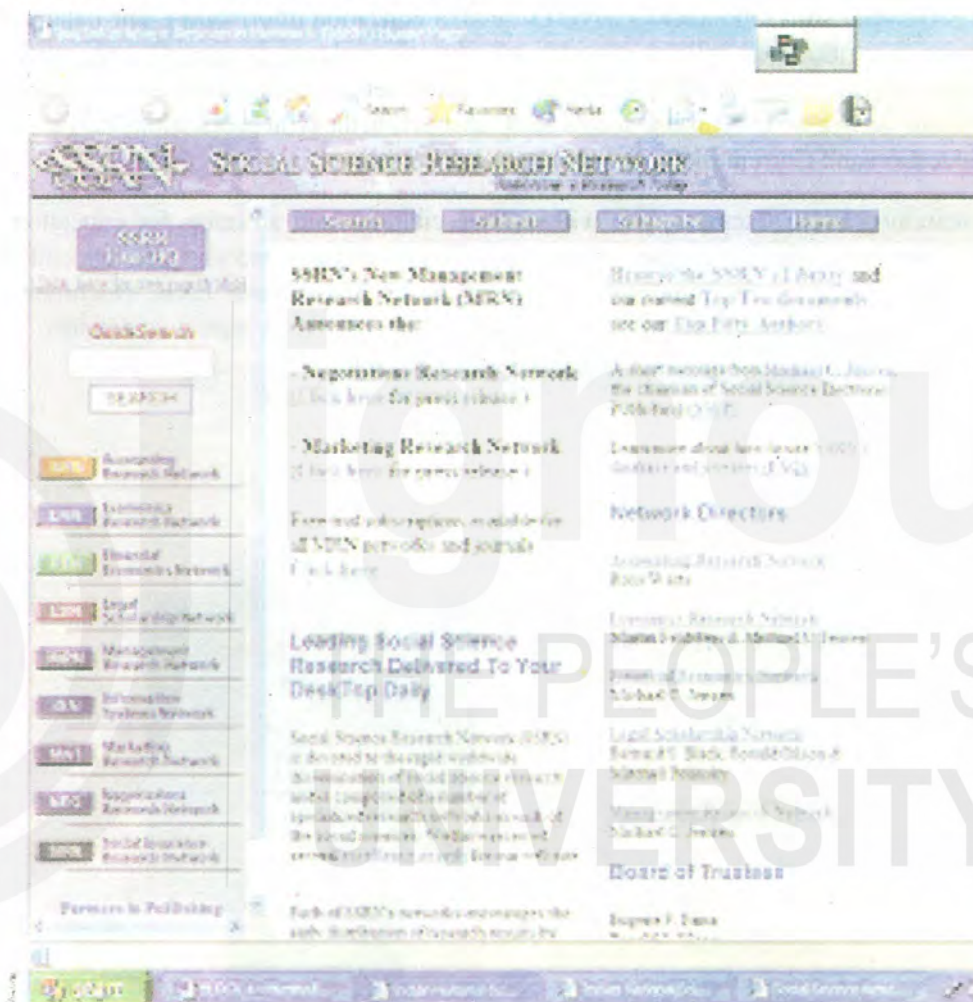


Fig. 15.17: Screen Snapshot of Social Science Research Network Website

iii) Humanities

The OCLC FirstSearch (<http://www.oclc.org/firstsearch/>)

The OCLC FirstSearch service connects a world of library users to a universe of information. FirstSearch gives library users instant online access to more than 70 databases, including the namely, OCLC databases: OCLC WorldCat, OCLC FirstSearch Electronic Collections Online, OCLC ArticleFirst, OCLC PAIS International, OCLC PapersFirst, OCLC ProceedingsFirst, and OCLC Union Lists of Periodicals. Best of all, library holdings are displayed up front, so users can easily identify items in their own library's collection. OCLC FirstSearch is a comprehensive and complete reference service with a rich collection of databases and with links to the World Wide Web, over 10 million online full-text articles, full-image articles from over 4,000 electronic journals, library holdings, and interlibrary loan. It supports research in a wide range of subject areas with well-known bibliographic and full-text databases in addition to ready-reference tools such as directories, almanacs and encyclopaedias.

Campus-Wide Information System is an information system intended to present an integrated view of the institution to the members of its community, as well as to alumnae, prospective students and others with an interest. Generally within the menu hierarchy, a broad array of local and Internet resources are made available. The CWIS is a growing resource always 'under construction'.

Campus Wide Information Systems are becoming an important resource base on the Internet for worldwide university campuses that are available online. Such resources provide an in-depth information about the desired campus — whether about academics, resources, course curriculum, library catalogues, databases and other library resources, campus accommodation, tuition fee, scholarships and the like. Besides the individual university campus-wide information systems, most of the universities have web pages for each department that provide contact details of the respective faculty and staff. There are a number of sites that aim to help you find web pages for particular University and also particular department of a university. Generally, the information that is included at such sites comprises the following:

- Information about student organisations and campus services
- Library catalogues and other databases
- Research opportunities
- University newsletters and journals
- Technical reports and preprints
- Administrative or academic department policies
- Schedules of lectures, plays or movies on campus
- Athletic event schedules
- Directory information
- Faculty research interests and publications
- Course offerings and syllabi

There are several advantages for the academic institutions in joining the system. These include the following:

- Making your department or organisation more visible
- Saving production time and expenses of distribution and printing
- Fast, easy updating of information
- Access library catalogues and other documents
- Accessible 24 hours a day, 365 days of year to all members of the community via the campus-wide network
- Cost-effective means to reach a world-wide audience
- Easy to use

i) STM

As an example, the homepage of Harvard University- Campus wide Information System (<http://www.harvard.edu>) is reproduced below:

HARVARD UNIVERSITY



SEARCH

INDEX

SITE MAP

HELP

HAPPENING AT HARVARD AUG 26, 2002



**From the literary page to the
public stage**

Summer School course connects
academics with action

[More](#)

**Harvard to mark anniversary of
Sept. 11 with solemn
remembrance**

President Summers will be principal
speaker [More](#)

**Kennedy School symposium to
examine public and media
response to Sept. 11**

Day of discussions set for Aug. 28
[More](#)

**Merit scholarships: Who is
really being served?**

Report questions whether merit
scholarships promote minority
access to higher education [More](#)

[More news](#)

[Events at Harvard](#)



latest research news
from Harvard
RESEARCH MATTERS

About Harvard

General information, maps, events

Academics

Links to Harvard's schools
Academic and research programs
Course catalogs, academic calendars

Admissions

Information about applying to
Harvard

Campus Life

Services for the University community
Student information, the arts, publica-
tions, athletics

Alumni

An online community for Harvard
alumni & alumnae

Directories

Telephone, mail, and email contact
information

Libraries, Museums

Research resources at the University

Info Technology

Computing resources at the Univer-
sity

Administration

Administrative offices
Resources for Harvard faculty & staff

Employment

Current non-faculty job openings

- **Visit the new President of Harvard, Lawrence H. Summers**
Inaugural speech, inauguration, biography, background info
- **Search** Harvard University web sites
- **Alphabetic index** of Harvard University web sites
- **Frequently** asked questions about Harvard
- **Contents** of this web site
- **Take an Online Tour of Harvard!**

Another example of CWIS is that of Massachusetts Institute of Technology (<http://web.mit.edu>). Broadly, the site provides you pointers to access detailed information regarding the following:

- Spotlights
- News—latest news, research, Open Courseware
- Academics—admissions, schools, courses, libraries
- Research—labs, centers, and programs
- Administration—offices and programs, giving to MIT
- Resources—for alumni, faculty, staff, and students
- Campus—life groups, activities, jobs available
- Events—calendar campus events and activities
- About MIT—facts, map, virtual tour, evolving campus
- Search—Anything you are interested in.

ii) Social Sciences

The following are some of the important websites on CWIS giving information about social science departments:

Campus Wide Information Systems (CWIS) Using WWW

(<http://www.hcc.hawaii.edu/hccinfo/cwis.html>)

Economics Departments, Institutes and Research Centers (EDIRC) in the World

(<http://ideas.uqam.ca/EDIRC/index.html>)

iii) Humanities

The following are some of the important websites on CWIS giving information on the humanities departments:

University of Massachusetts—English Department

(<http://www.umass.edu/english/links.html>)

Division of the Humanities, University of Chicago

(<http://humanities.uchicago.edu/humanities/>)

UCSC Humanities Division

(<http://humwww.ucsc.edu/>)

I) Technical Reports

Technical reports constitute an important information source for academics and researchers and for governmental agencies. Academics and researchers, particularly in STM disciplines often try to get access to technical/ research reports in the areas of their interest. Their importance lies in the following aspects:

- They provide additional material, which is usually not available in published literature.
- They provide details of interim/progress reports or completed R&D projects.
- They provide more details than papers in journals or conferences.
- They provide more nascent information than what is available in published literature.

With the help of Internet, such resources can be more easily accessed now. Examples include the following:

Langley Technical Reports Server (LTRS) (<http://techreports.larc.nasa.gov/ltrs/>)

LTRS is a service, which allows users to search available online NASA published documents, including meeting presentations, journal articles, conference proceedings, and technical reports. Many documents are available in compressed PostScript and PDF formats. All documents are unclassified and publicly available.

NASA Technical Report Server (NTRS) (<http://techreports.larc.nasa.gov/cgi-bin/NTRS>) NASA Technical Report Server (NTRS) is an experimental service that allows users to search the many different abstract and technical report servers maintained by various NASA centres and programmes. Specifically, it is a unified interface to many separate WAIS servers. NTRS is both a superset of the various servers, and a canonical listing of the servers. NTRS is intended for use by the various research communities targeted by the various report and abstract servers. It is open to all members of the Internet/World Wide Web community. A snapshot of its homepage is included below:



[Quick Start](#) / [More Info](#) / [Searching Tips](#) / [What's New](#) / [Give Feedback About NTRS](#)

Enter Search Keywords: (AND, OR & NOT are available)

Select the databases you would like to search:

- | | |
|---|--|
| ☐ Ames Research Center | ☒ Langley Research Center |
| ☐ Astronomy & Astrophysics (ADS) | ☐ LANL Astrophysics e-Prints (ADS) |
| ☒ Dryden Flight Research Center | ☒ Marshall Space Flight Center |
| ☐ Glenn Research Center | ☐ NACA Reports (abstracts only) |
| ☐ Goddard Institute for Space Studies | ☐ NACA Reports (full text) |
| ☐ Goddard Space Flight Center | ☐ Physics and Geophysics (ADS) |
| ☐ ICASE | ☒ CASITRS (RECONselect) |
| ☒ Jet Propulsion Laboratory | ☐ Selected Current Aerospace Notices |
| ☐ Johnson Space Center | ☐ Space Instrumentation (ADS) |
| ☐ Kennedy Space Center | ☐ Stennis Space Center |

m) **Library Catalogues**

The library catalogue lists all of the material available in the library with information on where to find it, whether it is available and how long you may borrow it for. Reservations can be placed on material that is currently on loan to another person. You can find the following materials on the library catalogue:

- Books
- Periodicals (e.g. journals, magazines, newspapers, reports and official publications)
- Pamphlets and Official Documents
- Theses
- Audiovisual material
- Microfiche and microforms
- Electronic resources

The Library of Congress Online Catalogue (<http://catalog.loc.gov/>)

The Library of Congress Online Catalogue is a database of records representing the vast collection of materials held by the Library. In addition to these records, the Online Catalogue provides cross-references, notes, and circulation status, as well as information about Library materials still in the acquisitions stage.

The Library of Congress Online Catalogue contains approximately 12 million records representing books, serials, computer files, manuscripts, cartographic materials, music, sound recordings, and visual materials. The Catalogue also displays searching aids for users, such as cross-references and scope notes. The catalogue records reside in a single integrated database; they are not separated according to type of material, language of material, date of cataloging, or processing/circulation status. As an integrated database, the Online Catalogue includes 3.2 million catalogue records from an earlier database. These catalogue records, primarily for books and serials cataloged between 1898 and 1980, are being edited to comply with current cataloging standards and to reflect contemporary language and usage. Placed below is the screen shot of the LC Online Catalogue web site.

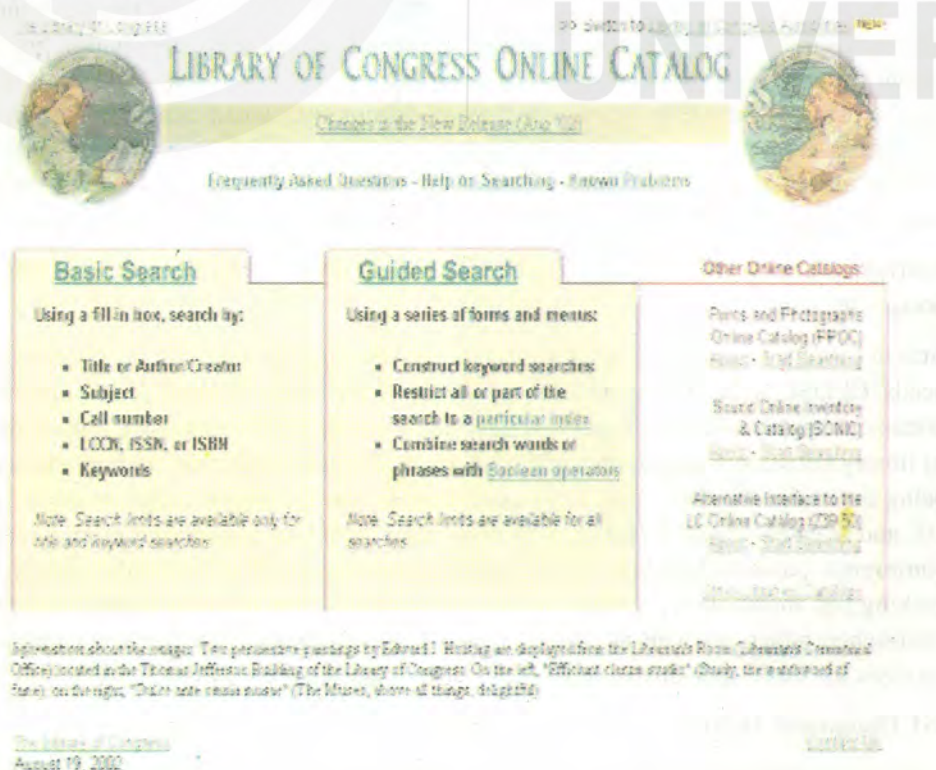


Fig. 15.19: Screen Snapshot of Website of Library of Congress Online Catalog

Questia (<http://www.questia.com/>)

Believed to be one of the world's largest online library. The documents are arranged subject wise basis.

n) **Patents**

According to U.S. Patent and Trademark Office publication "A patent for an invention is a grant of a property right by the Government to the inventor...The right conferred by the patent grant is...the right to exclude others from making, using, or selling the invention. What is granted is not the right to make, use, or sell, but the right to exclude others from making, using, or selling the invention."

There are also resources that give general information regarding patents, such as how to file a patent, etc. For example General Information Concerning Patents (<http://www.uspto.gov/web/offices/pac/doc/general/index.html>) is one such source.

There are volumes of such resources available on the Internet. Examples include the following:

US Patent and Trade Office - (<http://www.uspto.gov/>)

The PTO promotes industrial and technological progress in the United States and strengthens the national economy by:

- Administering the laws relating to patents and trademarks.
- Advising the Secretary of Commerce, the President of the United States, and the administration on patent, trademark, and copyright protection.
- Advising the Secretary of Commerce, the President of the United States, and the Administration on the trade-related aspects of intellectual property.

Provides free access to bibliographic data of US patents issued since 1.1.1976 and all has collection of full-text patents as well for US and also other countries.

o) **Document Delivery**

Document Delivery Service is an important component of the Library's interlibrary loan activity by which users' are facilitated not only to obtain information about the location of a document but the document itself. The Document Delivery Service or DDS is usually a fee-based service provided to faculty, staff and students, as well as to business, industry and individual researchers etc., by various universities, and other information providers/organisations. Document Delivery staff generally locates, photocopies, and sends directly to the customers office or home a variety of documents owned either by the library or has access to resources wherein the same are being accessed and sent through traditional post or electronically. There are a number of such services available on the Internet, most of which are fee based. Examples include the following:

British Library Document Supply Centre (BLDSC) (<http://www.bl.uk/services/document.html>)

British Library Document Supply Centre is a single largest source for all your document needs. BLDSC is the leading document provider in the world. A rapid and comprehensive document supply and interlibrary loan service from BLDSC's extensive collections and its library network to researchers and scholars in all kinds of libraries and organisations is being entertained by this centre. You can register for a range of services available for UK and overseas customers. If you want to order a copy of a particular journal article or conference paper straightaway, the Articles Direct service is probably what you are looking for. Inside web provides options for searching for relevant journal articles and conference papers as well as ordering them over the web. Facilities for several other services for more specialists copying of library materials are also included at this site.

ISI Document Delivery (<http://www.isinet.com/documentdelivery/>)

A document delivery service provides access to full-text items from virtually any publication within or outside of the ISI database. This flexible, fast, and convenient service provides

wide coverage, reliable customer support, fast and varied delivery options, and efficient management tools. The system offers the following:

- One-stop shopping — with comprehensive coverage that includes documents from research journals, conference proceedings and papers, book chapters, technical reports, government reports, annual reports, standards, and monographs.
- Helps to researchers to get the information they need fast - processing orders for documents in the ISI collection within 24 hours of receipt, and providing delivery options that include fax (30-minute delivery upon request), courier, or traditional mail service.
- Provides easy tracking of orders and accounts - via a professional customer support staff that helps verify order status and confirm order pricing and payment balance.

Eases record keeping by eliminating hidden costs: all materials have a standard processing fee, a variable copyright fee, and are copyright cleared.

p) Reference Sources,

A vast number of reference sources such as dictionaries, directories and other reference sources are freely accessible on the Internet. Reference sources have been categorised in two groups as listed below based on the nature of their coverage.

- Specific Reference Resources
- General Reference Resources
- **Specific Reference Resources**

Humanities Reference Sources (<http://www.library.siue.edu/lib/info/refhum.html>)

- **General Reference Resources**

Web of On-line Dictionaries: (<http://www.yourdictionary.com/>)

Links to on-line dictionaries in languages from Aklon to Zulu, as well as multilingual dictionaries and other language tools.



Fig. 15.21: Screen Snapshot of Dictionary.com Website

Dictionary.com: (<http://www.dictionary.reference.com>)

Dictionary.com is produced by Lexico Publishing Group (<http://www.lexico.com/>), a leading provider of language reference products and services on the Internet. This site provides links to various dictionaries, thereby facilitating search to various dictionaries at one place. In order to use the dictionary, you need to simply type a word in the blue search box that appears at the top of every page and then click the option 'Look it up' button. This will perform a search for the word in the several dictionaries hosted on this site (list of all dictionaries covered herein appears in the homepage of this site). If you don't know how to spell the word, just guess, you will get a list of suggestions if you are wrong.

Thesaurus.com: (<http://www.thesaurus.com>)

Thesaurus.com is also produced by Lexico Publishing Group. In order to use the thesaurus, simply type a word in the gold search box that you will find on the web page when you access this site and click the option 'Look it up' button. A list of synonyms and antonyms will be returned. The thesaurus that appears on this site is Roget's Interactive Thesaurus.

Encyclopedia.com: (<http://www.encyclopedia.com>)

Encyclopedia.com is the Internet's premiere free encyclopedia that provides users with more than 57,000 frequently updated articles from the Columbia Encyclopedia, Sixth Edition. Each article is enhanced with links to newspaper and magazine articles as well as pictures and maps — all provided by eLibrary. This eLibrary is a comprehensive digital archive for information seekers of all ages where one can search across 13 million documents from full-text newspaper and magazine articles, television and radio transcripts, international newswires, classic books, maps, photographs, as well as major works of literature, art and reference books. One can find both current and historical events within the diverse eLibrary archive.

Other Examples including general reference resources are as follows:

Refdesk.com (<http://www.refdesk.com>)

Refdesk is only about indexing quality Internet sites and assisting visitors in navigating these sites. Refdesk.com has three goals: (1) fast access, (2) intuitive and easy navigation and (3) comprehensive content. The site has extensive coverage in all the areas.

Encyclopaedia Britannica Online (<http://www.britannica.com>)

It used to cost up to £3,000 to purchase a complete set, but now the encyclopaedia's 32 volumes, or 44 million words, are accessible on the Internet - for free! What's more, it offers more than just access to entries in this famous encyclopaedia, and for each search topic you input, it provides links to some of the web's most informative sites, plus pointers to books and magazine articles that relate to your query.

Merck Manual (<http://www.merck.com/pubs/>)

The Merck Manual is the one of the most widely used medical texts in the world. Written by over 300 experts, it covers all but the most obscure disorders.

Yahoo! Reference (<http://dir.yahoo.com/Reference/>)

Yahoo! Reference covers - dictionary, encyclopaedia, quotations, and world facts.

World Fact Book (<http://www.odci.gov/cia/publications/factbook/index.html>)

Revised and updated every year, and compiled by the CIA, this is a good site for basic factual information on all the countries of the world: geography, people, government, economy, communications, transportation, and military. It also provides access to various maps.

The World Wide Gazetteer (<http://www.c-allen.dircon.co.uk>)

This site offers a substantial amount of practical information from country fact-files, drawing on a huge database, with links to maps, sources of information on the Web, current affairs, government and the economy, travel and flight data, and other resources.

q) Other Resources

Apart from the type of resources explained above, there are several other types of Internet resources which are grouped here as other resources.

Data Centres

Data Centres are primarily sites comprising databanks that provide actual statistical data related to specific subjects. These sources are of utmost importance in areas where statistical data is key input. Examples include the following:

The Humanities and Social Sciences Data Center

(<http://www.scc.rutgers.edu/datacenter/humanities/about.htm>)

This data centre at Alexander Library is a unique resource for Rutgers faculty, students, and other scholars. The Data Center facilitates access to the Rutgers University Libraries' Government, ICPSR (Inter-University Consortium for Political and Social Research). The Center serves as a clearinghouse for existing or newly created data collections located elsewhere.

The Data Center's holdings provide access to a number of full-text databases, such as the Packard Humanities Institute database, the Thesaurus Linguae Graecae, the Oxford English Dictionary, CETEDOC, and Letteratura Italiana Zanichelli. It also offers a convenient point of access to full-text Web-based databases, for which it provides supplementary documentation and tutorials. Some of these databases are the African American Poetry Database, the English Poetry Database, ARTFL, and the Dartmouth Dante Project.

Research Resources/ Research Projects

Research projects are an integral component of a resource-base in any subject area. Research resources/research projects provide a good starting point for researchers and academicians in any subject area to navigate what is broadly available and locate sources in their area of research. Such resources also provide links to tools that help in research work in the requisite subject area of interest. Examples include:

Education-Line

(<http://www.leeds.ac.uk/educol/>)

Education-line is a freely accessible database of the full text of conference papers, working papers and electronic literature, which supports educational research, policy and practice.

Social Science Research Resources

(<http://socsci.colorado.edu/POLSCI/RES/research.html>)

The site provides a pool of resources under the following four broad categories in various areas of social sciences: (1) Citing Internet Resources; (2) Online Courses and Guides; (3) Summer Programs in Data Analysis; (4) Social Science Research Centers.

Voice of the Shuttle (<http://vos.ucsb.edu>)

Possibly one of the most comprehensive sources for humanities research on the net. A large index to Internet resources, divided into headings, such as literature, culture, gender issues, minorities and religious studies, as well as more common humanities subjects. Developed and maintained by English Department, University of California Santa Barbara.

Asia Resources on the World Wide Web (<http://www.aasianst.org/asiawww.htm>)

Hosted by the Association for Asian Studies, an extensive list of online resources on Asia in general, East Asia, Southeast Asia and South Asia, together with links to journals/newspapers, dictionaries, libraries, videos and art.

(Research-It!) (<http://www.itools.com/research>)

Access to dictionaries and acronym converters. Look up words in many language tools; translate words to/from various languages; plus library, biographical, geographical and financial tools, and US and Canadian telephone numbers.

Virtual and Remote Experimental Laboratories

A shared virtual learning environment is a resource-based approach to the provision of learning materials. It addresses some of the pressures on higher education caused by the changing student population. Students will be able to access on-line learning resources and keep in touch with course mates from their home computers.

Discursive learning labs are very good for social sciences, languages, and humanities subjects. In the languages they are an effective way for students to practice their writing, speaking and listening skills among a peer group. They enable remote students to take part in the group discussion and arguments that are central to the humanities and social sciences. In fact, the collaboration tools that underlie discursive learning labs provide an excellent way for remote students to keep in touch with course-mates. Discussion boards should be adopted as a central component of a shared virtual learning environment. However, it may be indicated that remote laboratories without manipulation are effectively just video conferencing suites. Example include the following:

Web-Lab Information Page

(<http://weblab.badm.sc.edu/web-lab-information/web-lab-information.htm>)

Web-Lab is a Digital Library and Virtual Laboratory for Experimentation in the Social Sciences. Funded by the Digital Library Initiative at the National Science Foundation, it provides an online library of experimental software and data for research and teaching in Economics and Sociology. It also provides a network of collaborating laboratories for conducting experimental sessions. One needs to just click on the world map to connect to Web-Lab Col-laboratories around the world. A brief description of the Web-Lab project is available in pdf format. The site also facilitates pointers to link with other collaborating experimental sites in North America, Europe, and New Zealand. At the University of South Carolina there are three experimental laboratories: ExNet in Sociology, Clower-Lab in Economics and Beam-Lab in Business and Economics that can be accessed from this site.

Directories

Directories are search services, which involve human input in identifying relevant resources and allocating them to particular subject category or theme. Tools are usually searchable and browsable via the directory. Resources are not evaluated in terms of quality prior to their inclusion in the directory. There is galore of such search services available in Social Sciences via Internet. All types of directories are available be it specialists, e-mail directory, institution or the like. Examples include the following:

DARE: Directory in Social Sciences - Institutions, Specialists, and Periodicals

(<http://database.unesco.org/dare>)

Covers about 10,000 worldwide references to social science research and training institutions, specialists, documentation and information services, and social science periodicals; references to peace and human rights training and research institutions.

DRIS - Research Information Systems Worldwide

(<http://www.niwi.knaw.nl/en/oi/dris/toon>)

Information DRIS is a worldwide overview of current research information systems. The information systems covered by DRIS provide information on research organisations, researchers/scientists (expertise), or research projects/programmes. Some information systems may contain also publications, patents and/or conference papers. However, pure bibliographical systems are not included in this inventory. DRIS is offered free of charge.

Online Documents

Traditional textbooks are not commonly made freely available over the Internet. However, some academics have created their own online textbooks and these are often of a very high quality, with comprehensive coverage and excellent interactivity. Examples include:

The Web Center for Social Research Methods

(<http://www.socialresearchmethods.net/kb/index.htm>)

The Research Methods Knowledge Base is a comprehensive web-based textbook that addresses all of the topics in a typical introductory undergraduate or graduate course in social research methods. It covers the entire research process including: formulating research questions; sampling (probability and non-probability); measurement (surveys, scaling, qualitative, unobtrusive); research design (experimental and quasi-experimental); data analysis; and, writing the research paper. It also addresses the major theoretical and philosophical underpinnings of research including: the idea of validity in research; reliability of measures; and ethics. The Knowledge Base was designed to be different from the many typical commercially available research methods texts. It uses an informal, conversational style to engage both the newcomer and the more experienced student of research. A fully hyperlinked text that can be integrated easily into an existing course structure or used as a sourcebook for the experienced researcher who simply wants to browse.

Online Multimedia Projects and Exhibits

The multimedia projects and exhibits have special significance as a resource for humanities area. Multimedia basically means combination of audio, visuals, text, graphics, 3 D visuals, etc. all on a single platform. Hence, facilitating the virtual display of the product/resource and the like. Examples include of such multimedia resources include the following:

Alive TV (<http://www.ktca.org/alive/season12.html>)

Alive TV presents the work of artists whose "work speaks to us through its unflinching portrayal of challenges overcome through sheer human spirit".

The Electronic Academic Village (<http://jefferson.village.virginia.edu/home.html>)

An intriguing mixture of texts, hypermedia projects, technical and research reports, and other projects, which apply technology to the arts, humanities and social sciences. Developed by the Institute for Advanced Technology in the Humanities, University of Virginia.

Humanities-Interactive (http://www.humanities-interactive.org/a_base.html)

Presented by Texas Humanities Resource Center, Texas Council for the Humanities, this site presents about fifty web exhibitions that cover the scope of human civilisation and culture in graphical presentations. Other current exhibits include: "Border Studies"; "Bonfire of Liberties: Censorship of the Humanities"; "Newscast from the Past: June 15, 1215" MPEG movie clips.

Library of Congress Online Exhibits (<http://lcweb.loc.gov/exhibits/>)

The site provides an invaluable resource pertaining to art exhibits and provides access to exhibitions currently on display; Treasure-Talks; Exhibitions currently on tour and much more resources. A good site for artists and art lovers.

Others

Besides the above-indicated type of resources, there are several other resources as well. New resources are being added to the Internet on continuous basis. Hence, any list is liable to date.

It may However, be re-iterated that within different types, the sources fulfill more than one condition for being categorized under different types, Therefore, overlap of resources under different types is un-avoidable.

Information Resources by Resource Category

Resource categories are a way of categorising the different types of information that can be found on the Internet. These are basically service oriented, facilitating access to resources. The resources that can be grouped under this category are generally subject specific. The following are the main categories that can come under this, which are briefly outlined below with suitable examples for each case:

- a) Subject-based Gateway Services and Virtual Libraries/Resource Guides
- b) Subject Catalogues and Directories
- c) Rating and Reviewing Services
- d) Search Engines, Meta Search Engines, Search Tools and Search Tips
- e) Virtual Laboratories.

a) Subject-based Gateway Services and Virtual Libraries/Resource Guides

Generally developed by librarians or subject experts, these are an excellent source providing access to detailed descriptions of high quality resources. Some of these are known as virtual libraries; some are referred to as Subject-based Gateway Services while some others are called Resource Guides.

These resources are the most powerful resources in terms of the coverage and also authenticity. This can be the guideline for the navigators to look for the subject-based resources.

Subject gateways are online services and sites that provide searchable and browsable catalogues of Internet-based resources. Subject gateways typically focus on a related set of academic subject areas. The simplest type of subject gateways is sets of Web pages containing lists of links to resources. Some gateways index their lists of links and provide a simple search facility. More advanced gateways offer a much-enhanced service via a system consisting of a resource database and various indexes, which can be searched and/or browsed through a Web-based interface. Each entry in the database contains information about a network-based resource, such as a Web page, Web site, mailing list or document. Entries are usually created by a cataloguer manually identifying a suitable resource, describing the resource using a template, and submitting the template to the database for indexing. Subject gateways are also known as subject-based information gateways (SBIGs), subject-based gateways, subject index gateways, virtual libraries, clearing houses, subject trees, pathfinders and other variations thereof.

There are a considerable number of web-based gateways that can be used to locate network-based resources in a particular subject area. Nearly all of these gateways have unique features, additional subject-based services, and different approaches to how information about network-based resources is stored in the resource are also included. The basic characteristics of subject gateways are briefly given below:

Basic Gateway Facilities: Most subject gateways allow the end-user to either search or browse the databases of resource descriptions.

Additional Facilities: Some gateways provide extra facilities for enhanced searching. SOSIG, the Social Science Information Gateway, incorporates a thesaurus containing social science terminology. This gives users the option of generating alternative terms/keywords with which to search the resource catalogue. SOSIG also allows users to search on resources that are located in distinct geographic areas.

Additional Subject-Related Services: Some subject gateways have built considerable number of related services and information sources around their core searchable/browsable gateway. **Biz/ed**, (<http://www.bized.ac.uk/>) the Business and Economics information gateway for students, teachers and lecturers, contains large amounts of business and economic information, such as company financial data and economic datasets from sources such as the office of National Statistics.

Resource Cataloguing: The key difference between subject gateways and the popular automated large-scale web indexing systems such as AltaVista is the quality of the results, which the end-user receives.

There are excellent such Gateways in existence on Net in the area of STM, Social Sciences and Humanities. Some of these have been included here as examples:

i) **STM**

BIOME (<http://biome.ac.uk>)

A consortium-based hub (still in the process of development) providing access to quality resources on the Internet in the fields of agriculture, food, forestry, pharmaceutical sciences, medicine, nursing, dentistry, biological research, veterinary sciences, the natural world, botany, zoology, and more. It consists of five subject gateways, which are cross-searchable and cross-browsable. The service is coordinated by the University of Nottingham, who are joined by a formidable range of high-profile partners and content providers from the UK health and life science sectors.



Welcome to BIOME

Your guide to quality Internet resources in the health and life sciences

BIOME offers free access to a searchable catalogue of Internet sites and resources covering the health and life sciences.

Query: [Advanced Search](#) | [Help](#)

You can choose to search BIOME using the search box above for Internet resources in the whole of the health and life sciences, or choose one of the five subject-specific gateways from the menu bar to the left of the screen.

Unlike generic search engines, we only direct you to Internet resources that have been hand selected and quality evaluated.

We aim to provide a high quality service and so your feedback on this site is important us. Please spare a few minutes to complete our [feedback form](#).

PSIGate - Physical Sciences Information Gateway

(<http://www.psigate.ac.uk/newsite/about.html>)

PSIGate is the physical sciences hub of the Resource Discovery Network (RDN). PSIGate provides free access to high quality Internet resources for students, researchers and practitioners in the physical sciences, specifically in: astronomy, chemistry, earth sciences, physics, and science history and policy. Each resource in the main PSIGate catalogue has been selected by information professionals and subject specialists (mainly PSIGate staff and research postgraduates) to ensure relevance and quality.

JISC (Joint Information Systems Committee)

(<http://www.jisc.ac.uk/>)

The JISC is establishing posts known as Resource Guides in a number of subject areas. These posts will take over the maintenance of lists of resources known as Subject Guides. In addition, in co-operation with associated resource providers, the Resource Guides will provide information, documentation and training opportunities. Two such posts, covering the Arts and Humanities, and the Social Sciences, have been long established; others are being established, or have recently started their position. Subject Guides to JISC resources: Subject Guides are currently available in the following subject areas: Arts and Humanities, incorporating the subject areas of Language, Literature and Culture; Engineering, Mathematics and Computer Sciences, incorporating the Information Sciences; Aerospace and Defence Engineering; Life Sciences and Health Studies, including Biology and Biochemistry; the Natural World; Dentistry and Nursing Studies; Physical Sciences, including Chemistry and Physics; Social Sciences, including Business, Economics, Geography, Law, Politics and Psychology.

The JISC provides a number of electronic resources, services and development projects aimed to meet the needs of those working and studying within higher and further education. These are brought together here as Subject Guides under a number of subject headings. Each Resource Guide, when in post, takes over the maintenance of their associated Subject guide, and produces print versions and ephemera that complement the Web-based guides. There are overlaps within the subject areas and it is advisable to browse across the subjects. Each section of the Guides includes a list of resources with details of access arrangements. Some are free and require no license, subscription, or registration and can be accessed immediately. Other resources require subscription and/or registration. The JISC-funded Resource Discovery Network (<http://www.rdn.ac.uk/>) provides subject-based catalogues of descriptions of quality-assessed resources.

ii) Social Sciences

Examples of such gateways in the area of social sciences are given below:

Social Science Information Gateway (SOSIG) (<http://sosig.esrc.bris.ac.uk>)

The Social Science Information Gateway (SOSIG) is a freely available Internet service which aims to provide a trusted source of selected, high quality Internet information for students, academics, researchers and practitioners in the social sciences, business and law. It is part of the UK Resource Discovery Network. SOSIG is a kind of one stop shop for navigators looking for social science resources. Its several features include:

SOSIG Internet Catalogue: An online database of high quality Internet resources. It offers users the chance to read descriptions of resources available over the Internet and to access those resources directly. The catalogue points to thousands of resources, and each one has been selected and described by a librarian or academic. The catalogue is browsable or searchable by subject area.

Social Science Search Engine: This is a database of over 50,000 Social Science web pages. Whereas the resources found in the SOSIG Internet Catalogue have been selected by subject experts, those in the Social Science Search Engine have been collected by software called 'harvester' (similar mechanisms may be referred to as 'robots' or 'web crawlers'). All the pages collected stem from the main Internet catalogue. This provides the equivalent of a social science search engine.

Social Science Grapevine: This is the 'people oriented' side of SOSIG, offering a unique online source of career development opportunities for social science researchers in all sectors. Grapevine carries details of relevant training and development opportunities from employers and training providers. Researchers can also make their CVs available online which are freely accessible to all visitors to the site. Grapevine's Likeminds section provides a forum for exchange of ideas and information about potential research opportunities and partnerships. If you want to find contacts in your field, you can also check the social science departmental database.

My Account: This refers to your personal account on SOSIG. Use your portfolio; create your own personal view of the Web, using the high quality SOSIG Catalogue and the massive harvested "research-engine" database. You can find out which resources have been added recently in your area and register for regular email notification of important developments in your special area of interest. 'My Account' can also be used to post details of new social science conferences and courses to Grapevine.

SOSIG Catalogue is an excellent online catalogue of thousands of high-quality Internet resources relevant to social science education and research. Every resource has been selected and described by a librarian or subject specialist; search or browse the catalogue. Could be the first choice of navigator for finding social science resources. A snapshot of SOSIG Homepage is shown below:

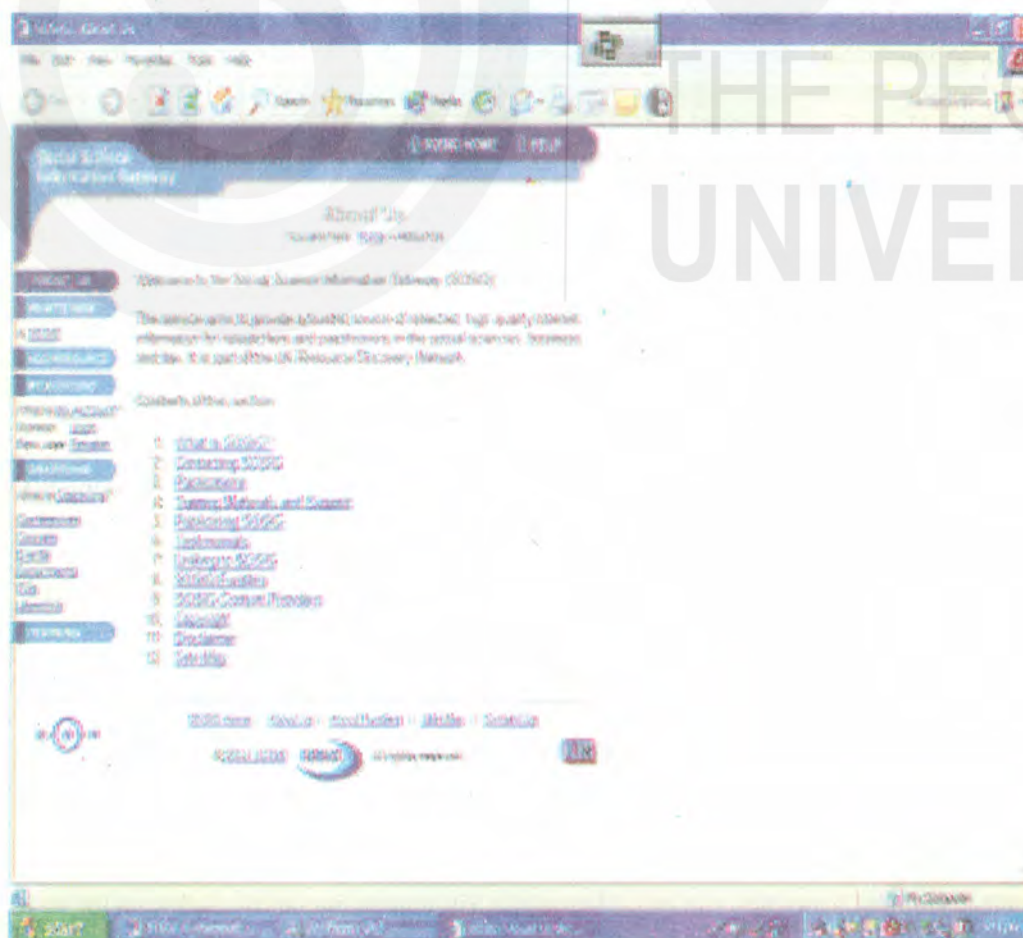


Fig. 15.23: Screen Snapshot of SOSIG Catalogue on the Website

Resource Guide to the Social Sciences (<http://www.jisc.ac.uk/subject/socsci/>)

The Resource Guide to the Social Sciences has been set up to raise awareness of the range of resources available and to offer a variety of activities to promote effective use of the resources. Each section of the Guide includes a list of resources with access descriptions. Some resources are freely available and can be accessed immediately. Others are conditionally free and/or require subscription and /or registration. Resources that are conditionally free or require subscription and/or registration may require a password that is issued by the subscription handling institution. A range of resources has been set up specifically to meet the needs of those working and studying in the social sciences. The guide focuses on resources funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Council of the UK, and the Economic and Social Research Council (ESRC). Resources covered are: bibliographic, references and research information; publications online; subject gateways; data services; datasets; spatial datasets; data visualization; software services and support for data processing; images, moving pictures and sound; learning and teaching; support services.

BUBL (<http://bubl.ac.uk/link/soc.html>)

This is a catalogue of 12,000 selected Internet resources and is an Internet-based information service for the UK higher education community. In Social Sciences alone it has 282 categories under which resources are listed. BUBL was the first national UK service to offer its users subject-based access to the Internet, through the BUBL subject tree initiative, which began in 1993. In the subject tree, resources are arranged together by subject area (i.e., all accountancy resources are located together, as are all geology, library and physics resources, etc.). The original Gopher-based subject tree was soon supplemented by a Web-based one, and both have now been incorporated into BUBL LINK (Libraries of Networked Knowledge). LINK contains thousands of links to Internet resources and services, and covers all main subject areas. Resource descriptions are searchable and subjects can be browsed by alphabetical order or Dewey Decimal Classification. BUBL subscribes to numerous mailing lists, which announce new resources and services on the Internet, giving their URL (Internet location) and a description of their content. Individuals also contact the service, sending information about a resource or service, along with its URL. BUBL staff evaluates these, and decide whether or not they are suitable for inclusion on the service.

iii) **Humanities**

Humbul Humanities Hub (<http://www.humbul.ac.uk/>)

It is one of the best sources in the Humanities covering large collection of high quality links to scholarly resources in the humanities. The Humbul Humanities Hub is a service of the Resource Discovery Network funded by the Joint Information Systems Committee and the Arts and Humanities Research Board, and is hosted by the University of Oxford. The Humbul Humanities Hub aims to be UK higher and further education's first choice for accessing online humanities resources. Provides pointers to resources on a wide dimension of areas in humanities.

Provides a number of services. One such service is My Humbul Include that brings Humbul's resources direct to a researcher or lecturer's web page. Through the use of dynamic html, users can embed a set of selected records within their own web page with the minimum of effort. Custom annotations can be added to supplement or replace Humbul's own descriptions. Meanwhile, Humbul maintains the metadata for each record - the link to the resource, the title, Humbul's description, the user's custom description and more - on behalf of My Humbul users.

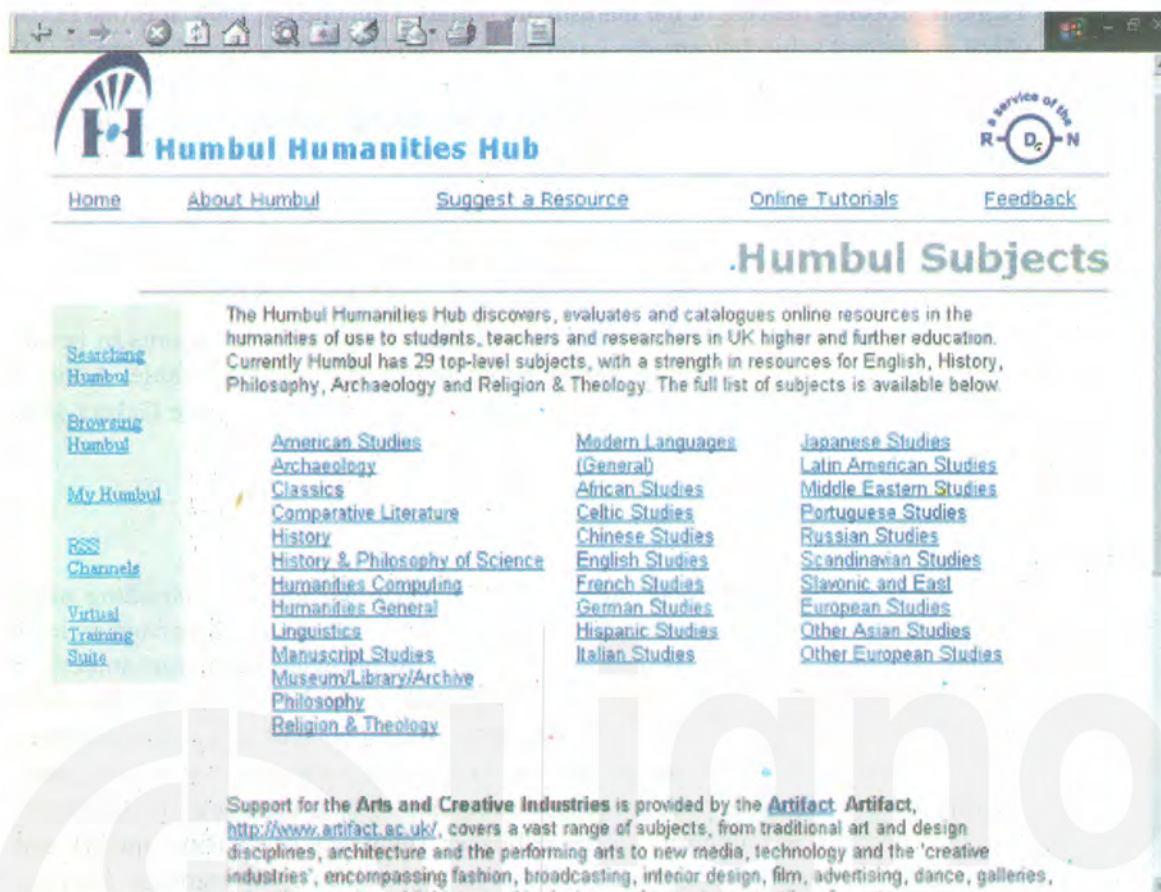


Fig. 15.24: Screen Snapshot of Humbul Humanities Hub Website

The WWW Virtual Library on Humanities: (<http://vlib.org/Humanities.html>)

This is another important subject gateway in the Humanities area, which facilitates the access to resources in various sub-areas under the main subject area of Humanities.

b) Subject Catalogues and Directories

Subject catalogues and directories perform similar functions, Hence, have been put together. They are similar to resource guides or resource catalogues with a subtle difference in terminology. It may be indicated that the subject catalogues and directories do not discriminate between sites in terms of their quality. Those involved in developing and maintaining the tools are concerned with the subject relevance of the materials and not necessarily with their quality —are involved in identifying potentially relevant resources and also in evaluating their quality.

Unlike search engines, subject catalogues or humans create directories. Sites re-submitted by their authors or identified by the site developers and then assigned to an appropriate subject category or categories by the catalogue maintainers. Thus, the basic resources are covered under subject listings or indexes in the resource directory. Subject catalogues tend to be both searchable and browsable through a hierarchy of subject headings. The best way to start is to investigate a general subject listing, like Yahoo or BUBL LINK (which is maintained in DDC order by librarian volunteers) or collections of specific guides to various resources, like those maintained by the Clearinghouse of Subject-Oriented Internet Resource Guide. Some popular Subject Catalogues and Directories are briefly described below:

Yahoo! (<http://www.yahoo.com>) and (<http://www.yahoo.co.in>)

Yahoo is probably the best of the mainstream portals. Launched in 1994, it is one of the oldest and largest subject directories covering over 750,000 websites divided into 25,000 categories. Yahoo is not a search engine, but a directory presented as a hierarchical subject index, which can be searched. It is a good starting point for searching for more broad or general rather than area-specific information. Also, it has good sections on reference resources, libraries, publishers, etc.

Galaxy (<http://www.galaxy.com/>)

Galaxy was initiated in 1993 and went live in early 1994, and also claims to be the oldest directory. It is similar to Yahoo being searchable and browsable. Authors submit their resources, which are then examined by a member of the Galaxy staff and allocated to an appropriate subject category or categories.

c) Rating and Reviewing Services

These services came into force during 1990s when trend towards providing more effective access to Internet resources through various forms of site ratings and reviews became prevalent. By using such services we get a list of reviewed Internet sources in respective subjects. This seems to be quite useful particularly to librarians while deciphering the worth of the site from information content point of view. These facilities tend to include materials of popular appeal and cover a wide range of subject areas. Sites included in these services are judged according to their accuracy, completeness, utility, credentials and authority of the publisher, frequency of revision, quality and effectiveness of presentation, elegance of design, and quality of the graphics. They are then awarded a star rating (as in case of eBLAST—Encyclopaedia Britannica's Internet Guide) and assigned to appropriate subject categories, while in some other cases such as Lycos rates top 5% sites numerically. The searcher could search any subject category of his/her interest and retrieve the information. Examples of rating services include the following:

Encyclopaedia Britannica's Eblast (<http://www.eblast.com/>)

A guide to the Web's top sites, chosen by Britannica's editors. Eblast cuts down on the frustration of Internet overload by narrowing choices to selected sites. Formerly known as Encyclopaedia Britannica Internet Guide, this recently redesigned hierarchical index to the web is uncluttered and easy to use. Britannica editors choose, annotate, and rate each site as well as provide lots of other information, at a glance, to help you decide if a link is worth the visit. It is a very useful site for students looking for quality research. Annotations generally include the title, author or responsible party, a star rating, and a concise description for the site. In addition, Site Statistics feature useful details such as site registration, number of visits, speed, and related topic links. The site's place in the Britannica subject hierarchy completes the annotation, providing links for further exploration.

Lycos Top 5% (<http://search.lycos.com/>)

Reviews sites in all subject areas. Provides link to (<http://www.alltheweb.com/>) for further confirmation on quality of ranked sites.

Argus Clearinghouse Rating System (<http://www.clearinghouse.net/ratings.html>)

The snapshot of this homepage of this site shown in Fig. 15.25 gives a clear picture about the site.

Ratings System

[Navigation](#)[Search/Browse](#)[Main Page](#)[FAQ](#)

How Guides are Rated

Guides that have been accepted by the Argus Clearinghouse are rated on the following five criteria:

[Level of Resource Description](#)[Level of Resource Evaluation](#)[Guide Design](#)[Guide Organizational Schemes](#)[Guide Meta-Information](#)

Guides receive a rating of 1 to 5 checks for each criterion. An overall score of ✓✓✓✓✓ to ✓✓✓✓✓ is obtained by averaging the guide's score on these five criteria. These criteria are subject to change; any changes will be publicized.

Level of Resource Description: Descriptive information provides users with an objective sense of what the Internet resources cover, including:

- Description of the resources' content (ranging from keywords to abstracts).
- Description of the traffic levels, level of moderation, features (e.g., digests) for mailing lists and Usenet newsgroups.
- Intended audience for the resources.
- Description of the update frequencies for resources
- Access instructions for the resources.
- Technical performance levels of the resources (i.e., a server is frequently down).

Level of Resource Evaluation: Evaluative information provides users with a subjective sense of the quality of the Internet resources, including:

- Quality of the content of resources (e.g. discussion in mailing lists and Usenet newsgroups, information in a Web site).
- Assessment of the resources' usability (e.g., document layout, readability, appropriate use of graphics, organization).
- Authority (e.g., reliability) of resource authors.

Guide Design: Quality guides balance aesthetics with usability:

- Images (are images attractive, do they support ease of navigation, do they load quickly).
- Layout (does the author make appropriate use of headers, mixed font sizes, and white space).
- Navigational aids (it is easy to find your way around, do you have a consistent sense of context or understanding of where you are in the guide at any given time).

Guide Organizational Schemes: Guides can be organized in one or more ways, including:

- By subject (e.g., sports can be broken up into soccer, racquetball, etc.)
- By format (e.g., mailing lists, Web sites, etc.).

- By audience (e.g., academic vs. general users).
- By chronology (for a history guide).
- By geography (for a guide to a region).
- By authors (for a literature guide).

Additional points will be assigned to guides with multiple organizational schemes.

Guide Meta-information: Meta-information is information about other information. In this case, meta-information describes the guide itself, and may include:

- Mission of the guide: why it was created, what it contains and what it leaves out.
- How the guide was researched and constructed.
- Information about the authors, their professional or institutional affiliations, and their knowledge of or experience with the subject.
- Information on how to contact the author and submit feedback and suggestions.
- Update frequency.

Exceptional Guides and Guide Design Techniques

Each month the Argus Clearinghouse staff chooses an exceptional guide as the Digital Librarian's Award recipient. Close examination of these guides will provide a wealth of information about superior guide design and organization.

▲ return to top

Fig. 15.25: Screen Snapshot of Argus Clearing House Ratings System Website

d) Search Engines, Meta Search Engines, Search Tools and Search Tips

As indicated earlier, the resources available via Internet are colossal in all fields of study. There may be instances one does not know or is not aware of any of the above resources or sites. Where does one go? The answer is search engines. Search engines are one of the primary ways of finding resources on the Internet. Also called spiders or crawlers, they constantly visit web sites on the Internet to catalogue web pages. However, it may be indicated that searching with search engines returns at times large number of hits, which is not feasible to go through. For more detailed information on search engines, see Unit 13 of Block 4 of this course. Some more examples are given below in addition to those given in that Unit.

Altavista (<http://altavista.digital.com>)

One of the most efficient search engines, which also translates the text of Web sites for you, in six languages. But do click on their 'Help for simple query' before you start your search and print out some of the information. Also offers a range of helpful 'Refine search options'.

Copernic (<http://www.copernic.com>)

This is another sophisticated Internet search application program that can be downloaded for free. It can query several search engines, directories, Usenet archives and e-mail address databases at once, providing access to some 30 major search engines and other information sources. It displays the results on a single page in your browser, stores them on your disk, organises them in a clear manner by relevance and score, and you can later retrieve them for offline browsing.

Debriefing (<http://www.debriefing.com>)

A good metasearch engine that finds the first 50 or so best matches to a search query (from Yahoo, Altavista, Excite and Hotbot) very quickly.

Google (<http://www.google.com>)

An excellent and innovative search engine that uses the number of links to a site to rate its importance. And 'I'm feeling lucky' button automatically takes you to the first Web page - i.e., the best result/match - returned for a query. Click on the bar graph at the beginning of the result to see which pages link to the particular page. Definitely one of the coolest search engines; strongly recommended.

Hotbot (<http://www.hotbot.com>)

Easy to use, fast and powerful, and one of the most consistently top-rated search engines. It offers a single form (in a pull-down menu) to search for a phrase, person, or any combination of search terms to fine-tune your output preferences. Good for 'exact phrase' searches, and also lets you search for sites with audio or video features.

ProFusion (<http://www.profusion.com>)

Metasearch engine; good for advanced searches.

Metacrawler (<http://www.metacrawler.com>)

Multiple-search resource which takes the best results from other search engines and directories and returns up to 30 hits from each site.

Bigfoot Directory Search (<http://www.bigfoot.com>)

Tracks down email addresses. Not infallible, but on the whole, works quite well.

Yahoo! People Search (<http://people.yahoo.com>)

Tracks down email addresses.

Examples of Search Tools and Search Tips

Searchability: Guides to Specialised Search Engines

(<http://www.searchability.com>)

A good descriptive guide to specialised search engines, and to search directories of search engines, evaluating their subject coverage and effectiveness.

Search Tools: A Guide

(<http://www.mmu.ac.uk/h-ss/dic/main/search.htm>)

An excellent guide to, and evaluations of, the major European Internet search tools, metasearch engines, email address finders, and various special search tools. Maintained by Richard Eskins at the Department of Information and Communications at Manchester Metropolitan University.

The Spider's Apprentice

(<http://www.monash.com/spidap.html>)

Advises on how to get the most from search engines, tips for search strategies, analysis of search engines, frequently asked questions, and more; this is a useful guide for beginners.

Finding Information on the Internet: A Tutorial

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

From the Library at the University of California at Berkeley, this is an excellent and easy-to-use tutorial on the latest and best search tools, and the best search strategies - progressing from simple searches to advanced searching - together with a general introduction to the Internet and the World Wide Web. In addition to recommending search strategies that are adaptive to specific topics, it provides detailed search instructions to

five major search engines (Google, Altavista Advanced Search, Infoseek, Northern Light, and FAST Search).

e) Virtual Laboratories

The rationale of the Virtual Laboratories is to offer around the world, irrespective of their local infrastructure, the virtual access to a high technology - world-class real laboratory. It aims to provide access to a network of renowned scientists and laboratories, offering research, technology and consultation. The specialty of such site is its interactive nature in which it needs to be accessed. Examples include the following:

Virtual Chemistry: (<http://neon.chem.ox.ac.uk/vrchemistry/>)

Virtual Chemistry is created and maintained by research students at the University of Oxford (UK), uses the latest multimedia technology for virtual experiments. Besides including a range of virtual experiments, the site also facilitates access to Real Player and has additional links to multimedia learning. A snapshot of the site is given below:

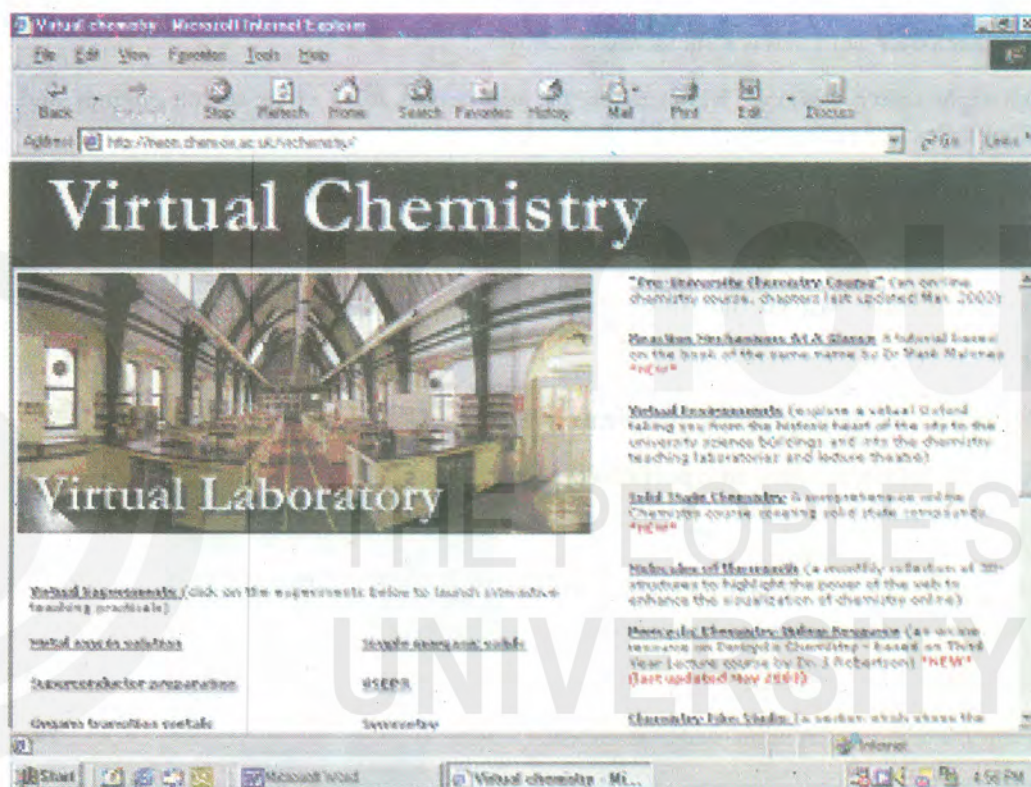


Fig. 15.26: Screen Snapshot of Virtual Chemistry Website

Information Resources by Organisation

Under this category, resources have been segregated on the basis of type of organisation, which is responsible for collating and providing access to the resources via Internet. These could be a collection of WWW page that are created and maintained by a particular organisation. Opening page is generally referred to as homepage. Organisational sites always include subject-based pages and personal homepages.

These resources could be from the following types of organisations:

- Booksellers, Publishers, Library Catalogues and the Media
- Commercial Online Information Retrieval Services
- Government, Government-Related and International Organisations Text Centres
- Professional / Research Associations
- Networking Organisations

f) University Departments

a) **Booksellers, Publishers, Library Catalogues and the Media**

As the name indicates, they provide a galore of information resources about publishing, bookselling and related aspects available on the Internet. The sources are briefly described below:

The World Wide Web Virtual Library: Publishers

(<http://archive.museophile.org/publishers>)

An extensive alphabetical list of Internet-accessible publishers from around the world. Also provides links to selected online bookstores and to broadcasters (listed by Country).

OCLC FirstSearch

(<http://www.oclc.org/services/>)

OCLC WorldCat (the OCLC Online Union Catalogue) For nearly three decades, libraries have shared their catalogues electronically to create the world's largest database of bibliographic information. WorldCat offers over 47 million bibliographic records-representing 400 languages-and holdings information vital for collection development, catalogueing, authority control, and retrospective conversion services. Through the OCLC FirstSearch service, users can access 70 databases including familiar names from leading information providers as well as resources provided exclusively by OCLC. OCLC databases include: WorldCat, ArticleFirst, Electronic Collections Online, PAIS International, PapersFirst, ProceedingsFirst, and the OCLC Union Lists of Periodicals.

b) **Commercial Online Information Retrieval Services**

These provide resources from commercial organisations available on the Net for information retrieval purposes. Access to such resources is generally fee-based. Example includes:

STN International

(<http://www.stn-international.de/>)

Online retrieval service provided by STN (Scientific and Technical Information Network), Germany covering over 100 scientific and technical databases with a strong emphasis on chemical information.

c) **Government, Government-Related and International Organisations and Text Centres**

These include resources put up by government, government-related and other international organisations. Generally such sources could be taken as authentic, since responsible agencies are answerable for information content. Such sites are good sources of information regarding government and other organisations. Examples include the following:

i) **STM**

FedWorld Informaion Network: (<http://www.fedworld.gov/>)

An information Server from NTIS (National Technical Information Service), USA. Includes US Government information servers, FTP, Gopher, and telnet services organised by subject. Also reports, databases and software from NTIS are also included here.

ii) **Social Sciences**

United Nations Economic and Social Council (<http://www.un.org/Overview/Organs/ecosoc.html>)

The Economic and Social Council was established by the Charter as the principal organ, under the authority of the UN General Assembly, to promote higher standards of living, full employment, and conditions of economic and social progress and development

worldwide. The site contains information on current UN development programmes as well as a selection of surveys of economic and welfare conditions throughout the world. Full-text of all UN's sessional documents from 1994 onwards are also available.

ICPSR (<http://www.icpsr.umich.edu/index-medium.html>)

The Inter-University Consortium for Political and Social Research (ICPSR), established in 1962, is an integral part of the infrastructure of social science research. ICPSR maintains and provides access to a vast archive of social science data for research and instruction, and offers training in quantitative methods to facilitate effective data use.

IASSIST (<http://datalib.library.ualberta.ca/iassist/>)

International Association for Social Science Information Services and Technology (IASSIST). IASSIST is an organisation dedicated to the issues and concerns of data librarians, data archivists, data producers, and data users. This unique professional association assists members in their support of social science research.

The Center for Basic Research in the Social Sciences (CBRSS)

(<http://www.cbrss.harvard.edu/index.htm>)

This Center was founded at Harvard University in June 1998 under the auspices of the Faculty of Arts and Sciences (FAS). The mission of this Center is to foster and improve basic social scientific research. It also supports seminars, workshops, and conferences, as well as a variety of teaching and training activities, including financial support for graduate dissertation research, for other Harvard research opportunities, and for implementation of innovative courses, besides other services. The site proves a good resource for those interested in basic research in the social sciences.

iii) **Humanities**

Some of the important institutions which provide access through Internet, to resources in the disciplines of Humanities are briefly discussed below:

UNESCO — Social and Human Sciences Documentation Centre

(<http://www.unesco.org/unescdi/>)

All UNESCO information sources are available from this unique access point. They include: 129 bibliographic, referral (directories, projects, etc.) and full-text databases produced by UNESCO in its domains of competence: education, natural sciences, culture, social and human sciences, communication and information. In addition, 58 information services located at UNESCO Headquarters and in the Field Offices can be accessed. Access, among others, to the full-texts of official UNESCO documents (about 20,000), the photobank (10,000 images), the worldwide translations database (1.3 million bibliographic records), thesauri, some links directories for specific themes (worldwide portals for libraries, archives, poetry, oceanography, etc.) is also available. These information sources are classified by type and by theme reflecting the main areas of activity of UNESCO.

The National Endowment for the Humanities (<http://www.neh.gov/>)

The NEH is a federal agency that supports learning in history, literature, philosophy, and other areas of the humanities. It funds research, education, museums, exhibitions, documentaries, preservation, and activities in the states.

National Humanities Institute (NHI) (<http://www.nhumanities.org/>)

NHI promotes research, publishing, and teaching in the humanities. The NHI website includes an online journal, NHI publications and research listings, an electronic bulletin board, and links to other humanities research on the net.

Text Centres

Text Centres provide the primary means of access to, as well as information on, full-text scholarly resources available at the various institutions. Texts are generally arranged by language, subject, and searching interface. Many such facilities do not point to digital facsimiles of texts; rather, they direct patrons to 'searchable' encoded texts and texts

where one finds requisite information. However, some provide pointers to the text itself. The Electronic Text Center supports the research and instructional use of electronic texts, mainly primary sources in the humanities and social sciences, from a dedicated facility.. At many such sites the assistance for how to use and create e-texts by providing consultation tips, instructions and point-of-use guides are also included.

Electronic Text Center-University of Virginia Library (<http://etext.lib.virginia.edu/>)

The homepage of this text center is self-explanatory and highlights as to what such sites offer. It is reproduced below:

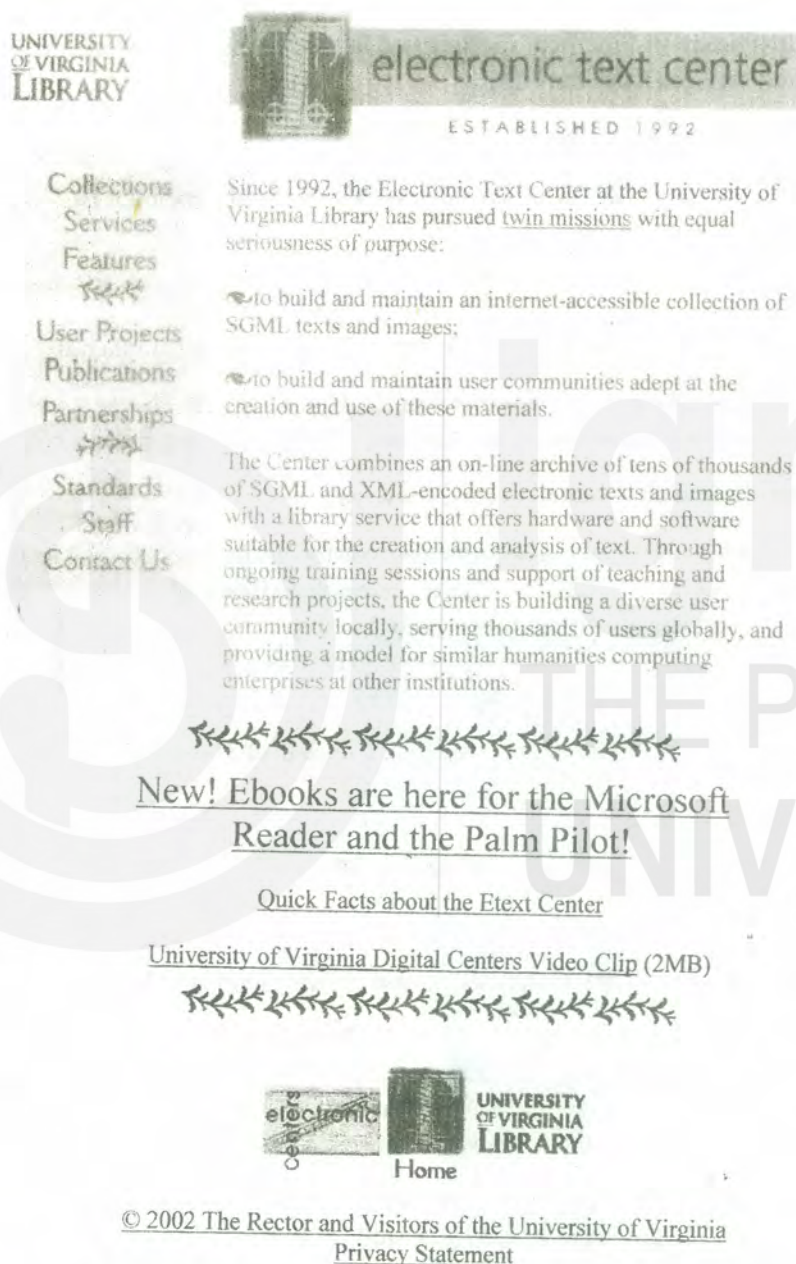


Fig. 15.27: Screen Snapshot of Electronic Text Center Website

Center for Electronic Texts in the Humanities (<http://www.ceth.rutgers.edu/>)

This Center is dedicated to helping people to access and implement research projects using electronic texts. The project is a joint effort at Princeton and Rutgers Universities.

d) Professional /Research Associations

Professional societies or associations are the general bodies of persons relating to the respective profession. These societies are set up with some societal missions with regard to the professions and hence they always work towards fulfilling the objectives for which these exist. Besides, such bodies also make constant efforts to design and develop new

mechanisms of looking at different things pertaining to the profession addressing core issues for the overall development of the field; finding solutions to the problems; networking of various related organisations; providing a platform for personal interaction with the coterie and the like. Therefore, resources which such organisations put up or provide access to, are very valuable and of contemporary nature. Such information servers provide information regarding professional associations like American Library Association, etc. Examples include the following:

i) STM

IEE Homepage (<http://www.iee.org/>)

Provided by Institution of Electrical Engineers (UK), covering professional, publishing and information services from the IEE.

BCS Net (<http://www.bcs.org./bcs/>)

British Computer Society (UK) is the provider of BCS Net and it includes BCS structure, events, awards, publications, etc.

ii) Social Sciences

American Council of Learned Societies (<http://www.acls.org/jshome.htm>)

A major coalition of scholarly organisations in the United States. The web site offers a linked list of constituent learned societies and links to other online scholarly resources.

Social Science Research Council (<http://www.ssrc.org/>)

The SSRC is an independent, non-governmental, not-for-profit, international association devoted to the advancement of interdisciplinary research in the social sciences. It sponsors interdisciplinary workshops and conferences, fellowships and grants, summer training institutes, scholarly exchanges, and publications. There is a useful hyperlink index of funders, affiliated institutes, and other organisations.

Social Sciences and Humanities Research Council of Canada

(<http://www.sshrc.ca/>)

Reports on grant support, conferences, projects, and current research.

iii) Humanities

Canadian Federation for the Humanities and Social Sciences (<http://www.fedcan.ca/>)

This is an association of scholarly societies in Canada. Created by an amalgamation of the former Canadian Federation for the Humanities (CFH) and the Social Science Federation of Canada (SSFC), the Federation came into being on April 1, 1996.

The Federation currently represents 69 learned societies, 71 universities and colleges and over 24,000 scholars and graduates active in the study of languages, sociology, literatures, religion, geography, psychology, anthropology, history, philosophy, classics, law, economics, education, as well as linguistics, women's issues, industrial relations and international development.

National Endowment for the Arts (<http://arts.endow.gov/>)

The National Endowment for the Arts provides national recognition and support to significant projects of artistic excellence, thus, preserving and enhancing USAs diverse cultural heritage. The Endowment was created by the Congress and established in 1965 as an independent agency of the federal government. This public investment in the nation's cultural life has resulted in both new and classic works of art reaching every corner of America.

National Endowment for the Humanities (NEH)

(<http://www.neh.fed.us/whoweare/index.html>)

NEH is an independent grant-making agency of the United States government dedicated to supporting research, education, preservation, and public programmes in the humanities.

The Association for Computers and the Humanities (<http://www.ach.org/>)

This Association is an international professional organisation. Since its establishment, it has been the major professional society for people working in computer-aided research in literature and language studies, history, philosophy, and other humanities disciplines, and especially research involving the manipulation and analysis of textual materials. The ACH is devoted to disseminating information among its members about work in the field of humanities computing, as well as encouraging the development and dissemination of significant textual and linguistic resources and software for scholarly research.

Research Associations

These associations are mainly concerned with the research and development activities and primarily focus on research and developmental issues. Examples include the following:

The Modern Humanities Research Association (MHRA)

(<http://www.mhra.org.uk/index.html/>)

Encouraging and promoting advanced study in the modern humanities, the Association aims to maintain the broader unity of humanistic scholarship in the face of increasing specialization, and attempts to fulfill this purpose especially through publications including journals, bibliographies, monographs, and other aids to scholarly research. Details of Association's publishing and funding activities are available from these pages. The site offers several options to various resources that are displayed below in the Menu of this site.

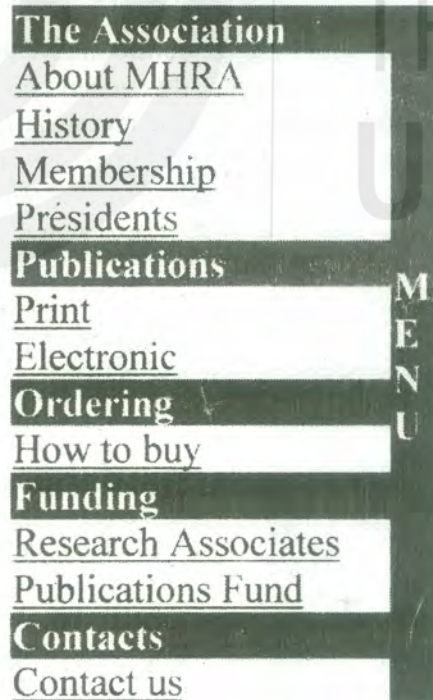


Fig. 15.28: Screen Snapshot of The Menu of Website of Modern Humanities & Research Associations

e) Networking Organisations

Generally related to network issues, these are good sources for up-to-date information on network developments. Examples include the following:

i) **STM**

IETF- Internet Engineering Task Force (<http://www.ietf.cnri.reston.va.us/>)

Working under the auspices of the Internet Society, the IETF provides management standards and technical information and many sources of further information.

TERENA- Trans-European Research and Education Networking Association

(<http://www.terena.nl/>)

The TERENA Secretariat (The Netherlands and Europe) gives information about TERENA and its predecessor EARN. Also gives European-wide networking initiatives including reports, calls for proposals, conference papers and abstracts. It also contains documentation for major Internet tools and links to international networking centres, e.g., InterNic.

(ii) **Social Sciences**

Social Science Research Network (SSRN) (<http://www.ssrn.com/index.html>)

SSRN is devoted to the rapid worldwide dissemination of social science research and is composed of a number of specialized research networks in each of the social sciences. These networks are:

Economics Research Network (<http://www.ssrn.com/ern/index.html>)

Financial Economics Network (<http://www.ssrn.com/fen/index.html>)

Legal Scholarship Network (<http://www.ssrn.com/lsn/index.html>)

Management Research Network (<http://www.ssrn.com/mrn/index.html>)

iii) **Humanities**

Coalition for Networked Information (CNI) (<http://www.ninch.cni.org/>)

CNI is an organisation dedicated to supporting the transformative promise of networked information technology for the advancement of scholarly communication and the enrichment of intellectual productivity. Some 200 institutions representing higher education, publishing, network and telecommunications, information technology, and libraries and library organizations make up CNI's Members. CNI and Dartmouth College announce the availability of a new web site, Collaborative Facilities (<http://www.dartmouth.edu/~collab/index.html>) designed to collect, organise, and disseminate information about model 'collaborative facilities' on college and university campuses. Visitors may 'tour' and analyse documents from facilities ranging from information commons to distance-education offices to centers that assist faculty in integrating teaching and new technology.

f) **University Departments**

University Departments are also an excellent source for information resources in various subject areas. This category has already been discussed under campus wide information systems/university departments.

Self Check Exercise

- 4) What do you understand by Data Archives? How do they facilitate as a resource base? Identify the drawbacks, if any.
- 5) The rationale behind Virtual Laboratories is to provide globally virtual access. Comment on this statement? Include suitable examples.
- 6) What are preprints and working papers? Give suitable examples.
- 7) How does document delivery service systems help a librarian or a researcher? Name the world's largest document delivery centre and briefly mention its services.

- 8) Briefly mention the nature and characteristics of the Subject-based Gateway Services. Name the important social science and humanities gateway services.
- 9) How are On-Line Writing Guides/Labs useful as a resource-base? Cite an example of such resources?

Note: i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of the Unit

.....

.....

.....

.....

.....

.....

15.4 SEARCHING THE INTERNET: WHERE TO START

Having known various search tools and facilities (explained in Unit 13 and 14 of this course), the searcher/navigator needs to know where to start in order to economise on time and get the best. The following gives a gist of where to start for accessing information on the Net.

- The first task when you are looking for quality information is to locate a site or service, which provides access to high-quality materials within the subject area of interest. You may look for something, which is stored in specialised searchable databases (invisible web). As it has been estimated that only 1/500th of the web is visible to search engines. So, to begin with subject gateways or portals may be a good option.
- Subject-based gateway services and virtual libraries would probably be the best starting point from where you can reach to the enormous number of guides. Using a gateway or a virtual library as a starting point can save you a lot of time and effort because you need not sift through thousands of outdated and useless sites. Facilities like Argus Clearinghouse can act as a great source to begin with.
- However, the searcher might not be in a position to either find a guide covering the requisite subject area or even if he/she locates a guide, there may be nothing of interest in it. The second step thus, should be to broaden the search. For instance, if you are looking for 'Breast Cancer in Women', you may look for a broader term 'Cancer'.
- If such approach does not prove useful, then one should move to general subject directories or catalogues, although they tend to be larger but less selective in terms of site quality.
- If still nothing useful is found, then the next option should be the search engine. These yield comprehensive search results, though these could be less specific.
- The final option would be to opt for a meta search engine, and still if nothing worthwhile is found, it may be time to presume that no relevant item is available on the Internet on that specific query, which is of course very rare.
- In such situations when you have exhausted options, you may want to ask someone through a discussion group.

The following pictorial representation gives a gist of main advantages and disadvantages of various search tools and facilities as summarised by Cooke, Alison in 'A Guide to Finding Quality Information on the Internet', which has been modified here by adding new dimensions. In general, the more the coverage is, the less will be the quality of the retrieved resources.

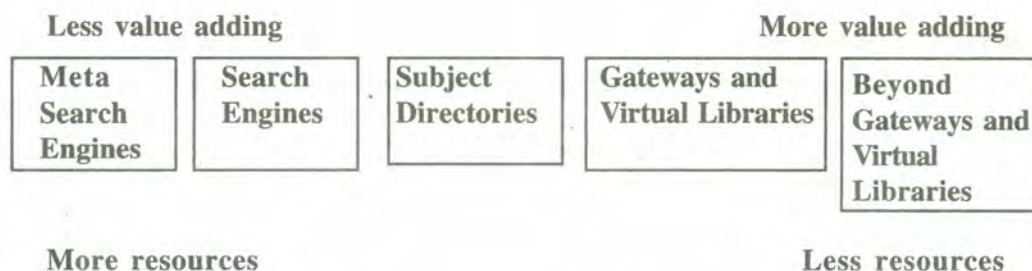


Fig. 15.29: Advantages and Disadvantages Search Tools

15.5 HOW-TO KEEP UP-TO-DATE WITH NEW INTERNET RESOURCES

It is important for academicians and researchers undertaking research to keep up-to-date with the latest literature in their fields. The wealth of information available to us is enormous. New resources are being added to Internet every now and then. Therefore, looking for a specific piece of information on the Internet is like searching for a needle in a haystack. How does one know which is the best site from where one can get information that one is looking for? You need sources to help make this task easier. Search tools, resource directories and current awareness services offer some solution. The Internet can be useful for maintaining current awareness within a particular subject field. It is important to keep a track of new sites that are added in the concerned area of specialisation. Keeping up-to-date is essentially a matter of following appropriate discussion lists and Usenet newsgroups. Other people recommend or announce new resources, which they find on the Net or have used. Similarly you can bookmark or preserve URL of a source in a file for future use, which you think, is appropriate. However, there are some sources available on the Internet, which keeps you posted about new resources. Examples include the following:

Internet Resources Newsletter (<http://www.hw.ac.uk/libwww/irn/irn.html>)

Internet Resources Newsletter is a free monthly electronic newsletter, edited by Heriot-Watt University Library staff and published by Heriot-Watt University. This free, monthly newsletter is very useful for academics, students, engineers, scientists and social scientists. Possibly, the most useful round up of new resources for academic users.

Scout Report (<http://scout.wisc.edu/>)

A weekly publication describing new resources of interest to researchers and educators and anyone else with an interest in high-quality online material. A very useful, regularly updated source of information for new resources covering all types of resources like full-text papers; tables of contents for new journals; forthcoming conferences; statistics and much more. The Report is available both on the web site, and in email form via mailing list subscriptions.

SENN- Scientific and Engineering Network News (<http://www.senn.com/>)

Includes top searches in various areas of science and engineering.

WWW Social Sciences Newsletter (<http://www.clas.ufl.edu/users/gthursby/socsci>)

This online newsletter (ISSN 1323-9376) is a part of the Social Sciences WWW Virtual Library and is provided as a service to the World Wide Web community. To receive the listings of new or improved WWW sites in the form of e-mail, you may subscribe to such

a service by sending an e-mail message to majordomo@clas.ufl.edu with the following information in the body of the message: subscribe socsci-news, Firstname Lastname, email address.

Netsurfer Digest (<http://www.netsurf.com/nsd/>)

Netsurfer is now charging money for full access to their content. However, some stuff is still free. For subscription to such a service, full details about the new subscription system can be found at: SUBSCRIPTION INFO: <http://www.netsurf.com/signup.html>.

BUBL LINK Updates (<http://bubl.ac.uk/link/updates/current.html>)

Selected Internet resources covering all academic subject areas. One could select appropriate subject area such as social sciences or Humanities and find out new resources.

Infomine : scholarly Internet resource collections (<http://infomine.ucr.edu/>)

Click on what's New next to each subject category to view new additions to the directory. You can also join an Alert Service, to receive e-mail information of new sites as they are added to the directory.

New Resource Alert Service lists new resources. The e-mail notification service can keep you informed of new resources as they are added to I-NFOMINE. You can initially choose (and later modify) one or more general subject areas of INFOMINE to receive alerts as well as the frequency with which they are received.

Academic Info - What's New (<http://www.academicinfo.net/new.html>)

Sign Up <mailto:madin@academicinfo.net> to receive their monthly announcement list.

Self Check Exercise

- 10) Distinguish between Search tools and clearly bring out their advantages and disadvantages.
- 11) How can you remain up-to-date with new Internet resources? Mention a few resources that can help you in this regard.

- Note:** i) Write your answers in the space given below.
ii) Check your answers with the answers given at the end of the Unit

.....

.....

.....

.....

.....

.....

15.6 SUMMARY

There are many types of information resources available, often more than we know what to do with them. Information flows in different ways in different subject areas. Knowing how information flows and becomes accessible helps us in deciding what is that we are looking for. The use of electronic media to publish information has significantly speeded up the information cycle, that is, information is available to the general public more quickly than it ever has been. There is a galore of information available via Internet for education, research, recreation and leisure purposes. There are many types of information resources available on every conceivable item via Internet. There are quantitatively mind-boggling

resources available in all areas of education, research and development. Be it Humanities, Social Sciences, Science, Technology and Medicine or Multidisciplinary areas, etc. The resources are free or fee based; both ephemeral and quality resources; short lived or permanent; standard or not standard; by individuals or organisations and the like.

Internet information resources include both documents and non-documents in electronic or e-format that provide information or a pointer to the information and can be accessed via Internet. These electronic resources could be of varied nature and can broadly be categorised as - Primary Sources of Information, Databases, Data sets and Other collections, Electronic Books, Reference Sources, Organisations and People, Meta resources, etc. This Unit describes important players in the e-information publishing and major ways of subscribing to e-resources. In this Unit, attempt has been made to give an exposure to the types of resources available in the various subject areas. Since such resources provides pointers to all forms of resources, thereby acting as a one-stop shop for most of user's information requirements. The Unit deals with the Internet information resources of interest to Libraries, besides specific resources in the area of Science, Technology and Medicine; Social Sciences; and Humanities. This list by no means is a comprehensive list, which is neither possible nor feasible. This is due to the fact that every hour new resources are being added to the existing pool of Internet resources. In the process some sites vanish as well. Therefore, any list is liable to date. Internet resources are growing at an unprecedented rate on every conceivable topic or subject. Its major strength lies in the huge diversity of material available aimed at all subjects, all levels and for all types of users. It offers a staggering range of data, information and software covering almost every imaginable subject. What one needs is to know the structure of these resources? The Unit covers various Internet resources, which have been categorized under three broad categories in order to make segregation of resources meaningful.

These three categories are — by resource type, by resource category, and thirdly by organisation type. In the first category sources have been divided on the basis of their form - for instance journals, table of contents, data archives, electronic journals, books and library collections, indexing and abstracting service resources, subject databases and the like. These resources provide a rich source of information in the said form of sources. For instance, when you are looking for journal collection in Humanities area, the resources listed under source type journals would facilitate you in identifying journals and related information that is available on the Internet instantaneously. Under the category or resource type are listed resources of prime importance for the navigators interested in research activity in any given area of interest or looking for any type of information in a particular area. It may however, be mentioned here that the resources listed here are a storehouse of various forms of resources under one umbrella. Special mention may be made of the subject gateways. Various resources that have been mentioned under this category are important from the point of view that any naive user/navigator who is looking for some information in the given area, these resources shall provide a good starting point and will definitely facilitate in locating some information that the user is looking for. Therefore, resources listed under this category are very important from users point of view. The third group comprises of the resources by organisation type. Here in, the organisations have been classified on the basis of their characteristics and the resources have been listed as examples. For instance, the organisations have been classified as — International Organisations; Text Centers; Professional Societies/Associations; Networking Organisations; University Departments and the like. The resources listed in here or the resources to which these sites provide pointers are undoubtedly authentic, quality and researched resources, since such organisations or centres have been set-up with a specific mission and are exclusively devoted to research and development in the specified areas. The Unit also provides an insight into such aspects pertaining to keeping oneself up-to-date with new websites and resources. The ultimate purpose of this Unit is to facilitate you to navigate the Internet resources successfully in any given subject area.

15.7 ANSWERS TO SELF CHECK EXERCISES

- 1) Internet information resources can be defined as resources that include both documents and non-documents in electronic or e-format that provide information or a pointer to the information and can be accessed via Internet.

These electronic resources could be of varied nature. Broadly, we could categorize them as follows:-

Primary Sources of Information: Such as electronic conference papers; electronic journals; electronic pre-prints and e-prints; electronic theses and dissertations; patents; standards; technical reports; project reports including status reports of current ongoing projects; News; software courseware; tutorials, Manuals and the like.

Databases, Data sets and Other collections: Such as abstracting and indexing databases; digital collections comprising images, audio, video; scientific data sets comprising numeric, property, structural databases; library catalogues; virtual libraries; museums and archives, etc.

Electronic Books: such as NetLibrary (<http://www.netlibrary.com/>); Ebrary (<http://www.ebrary.com/>), etc.

Reference Sources: such as dictionaries; encyclopedia; biographies; handbooks; thesauri and the like.

Organisations and People: Information about organisations and people ranging from funding agencies to libraries; information centres; research institutes; and experts; directories of people of varied nature (scientists; archaeologists, etc).

Meta resources: Resources that facilitate easier access to network based resources in a defined subject area and a plethora of such resources under various names are out there on the Internet, such as subject gateways; virtual libraries; clearing house; pathfinders and the like.

- 2) There is a mixed bunch of players in the market who publish electronic or e-information resources. These include traditional players offering electronic versions of their resources as well as several new enterprises offering new products and services that are 'born digital'. Besides, there are others also, such as subscription agents in their new role in the e-environment as electronic aggregators; institutions of higher learning and the like. We may broadly classify the Internet information resources publishers in the following categories:

Traditional Commercial Publishers: such as Elsevier Sciences' Science Direct (<http://www.sciencedirect.com/>);

Scholarly Societies/Bodies: Several such agencies are making electronic versions of their publications available online through their websites. For instance, Indian National Science Academy (INSA) scholarly publications are available online for free and unlimited access at <http://www.insa.ac.in>

Institutions of Higher Learning: Several institutions, academic and research, host their specialised collections like courseware, thesis and dissertations, etc., and also maintain subject gateways and portals; etc.

Electronic Aggregators: These facilitate intermediary services of aggregating electronic journals and other resources from different publishers and offer them to their clients through a single search system. Examples of aggregators are: JSTOR; OCLC; Lexis-Nexis and the like.

Others: Among the others are non-commercial publishers; individuals; etc. Many individuals, groups and non-commercial publishers are compiling resources dedicated to providing access to e-resources (generally offered free) on the Internet. Examples

of such meta resources include: Librarians' Index to the Internet (<http://lii.org/>); Vlib: The Virtual Library (<http://www.vlib.org/>); etc.

- 3) Journals are made available through the web using various modes and at varying price models. Some of the common ones are as follows:

Electronic Subscription bundled with Print Subscription: In such cases electronic subscription to journals is linked with their print versions. This means, e-version may be either offered free with their print counterparts; or a fixed token percentage may be charged on and above the print subscription charges or some other similar model.

Electronic Subscription with Campus Licenses: wherein unlimited access to subscribed journals is facilitated campus-wide.

Electronic Only: In such situations publishers are offering electronic versions only with a token discount.

Pay-per-look: Identification of e-articles and paying for such articles in full text on a per-look basis. This could be considered somewhat similar to print versions where articles of interest are identified and requested on payment basis via a document delivery system.

Consortium Licensing: with the popularity of consortium mode of subscription, publishers have started working out best pricing models suitable for different nature of consortia, whether at regional, local or national level.

- 4) A data archive is a permanent, electronic collection of datasets with accompanying metadata such that users of the data can acquire, understand, and use the data. Data archives are resource centres for analysts who use data for research and teaching. Data archiving is a method of conserving very expensive resources and ensuring that their research potential is fully exploited. Archives ensure that when technology changes, the data in their holdings are technically transformed to remain readable in the new environment. Their functions usually include: being more than a long-term backup, being more than an index or catalogue with pointers to datasets stored elsewhere, ensuring that data are preserved against technological obsolescence and physical damage, cataloguing their technical and substantive properties for information and retrieval, supplying them in an appropriate form to secondary users. Data archives have been established in most European countries and in the United States. They are actively used for testing hypotheses and for other scholarly purposes.

Advantages: The advantage of such resources also lies in the fact that while the libraries are constantly under pressure for want of more space, these service make up for that. Since they maintain the archives at their end while facilitating you to access the archives generally against fee. Another important point is that the library really does not have to worry about maintenance of the documents against any odds. Thirdly, the access to such archives is available round the clock. These are more useful in the areas of social sciences and humanities.

Drawbacks: There are some difficulties as well, primarily, if you are discontinuing buying access to such services, your permission to access the requisite archive ceases to exist. What happens to the subscription that you have paid for a specified time period? While in printed document you still hold the document for the years for which you have paid the subscription and acquired the document. However, in the online set-up many models are now being proposed and marketed by the data archiving and marketing agencies that vary from agency to agency, which needs attention.

- 5) The rationale of the Virtual Laboratories is to offer around the world, irrespective of their local infrastructure, virtual access to a high technology - world-class real

laboratory. It aims to provide access to a network of renowned scientists and laboratories, offering research, technology and consultation. The specialty of such site is its interactive nature in which it is accessed.

Examples of this type are as follows:

Virtual Chemistry: (<http://neon.chem.ox.ac.uk/vrchemistry/>)

Virtual Chemistry is created and maintained by research students at the University of Oxford (UK), which uses the latest multimedia technology for virtual experiments. Besides including a range of virtual experiments, the site also facilitates access to RealPlayer and has additional links to multimedia learning.

Virtual Laboratories in Probability and Statistics: (<http://www.math.uah.edu/>)

Created by University of Alabama's Mathematical Sciences Department, the project is partially funded by National Science Foundation (NSF). The project's focus is to provide interactive, web-based modules for students and teachers of probability and statistics.

- 6) The Internet is increasingly being used by academics to publish the full text of conference papers, draft papers or work-in-progress, and other similar material, often to facilitate peer review process. These are called pre-prints or working papers. Examples include the following:

London Business School, Centre for Marketing Working Papers (http://www.lbs.ac.uk/marketing/Working_Papers/working_papers.html) provides a list of working papers from the Centre. Available from 1995, it includes abstracts for all the papers and full text for most of the papers in Adobe Acrobat (PDF) format.

- 7) Document Delivery Services are an important component of the Library's inter library loan activity. Such systems do not only point to location of a given document in a particular library/information centre but also facilitates access to the document itself. There are several such initiatives available on the WWW that facilitate document delivery services in online mode that were earlier handled by the traditional manual means. These include the following:

British Library Document Supply Centre (BLDSC) (<http://www.bl.uk/services/document.html>)

BLDSC is the leading document provider in the world. A rapid and comprehensive document supply and interlibrary loan service from BLDSC's extensive collections to researchers and scholars in all kinds of libraries and organizations is being entertained by this Centre. The British Library has a large Document Supply Centre dedicated to the supply of copies of journal articles, books and other material. You can register for a range of services for UK and overseas customers. If you want to order a copy of a particular journal article or conference paper straightaway, the Articles Direct service is probably what you are looking for. Inside web provides options for searching for relevant journal articles and conference papers as well as ordering them over the web. Facilities for several other services for more specialists copying of library materials are also included at this site.

- 8) A collection of searchable (and usually browsable) resource descriptions which have been developed by library professionals and/or subject experts with the explicit aim of providing access to high quality sources of information is called a subject-based gateway source. The descriptions relate only to high-quality materials available via the Internet within a particular subject area.

The simplest types of subject gateways are sets of web pages containing lists of links to resources. Some gateways index their lists of links and provide a simple

search facility. More advanced gateways offer a much-enhanced service via a system consisting of a resource database and various indexes, which can be searched and/or browsed through a web-based interface. Each entry in the database contains information about a network-based resource, such as a web page, website, mailing list or document. Entries are usually created by a cataloguer manually identifying a suitable resource, describing the resource using a template, and submitting the template to the database for indexing.

The basic characteristics of subject gateways are as follows:

Basic Gateway Facilities: Most subject gateways allow the end-user to either search or browse the database of resource descriptions.

Additional Facilities: Some gateways provide extra facilities for enhanced searching. **SOSIG**, the Social Science Information Gateway, incorporates a thesaurus containing social science terminology. This gives users the option of generating alternative terms/keywords with which to search the resource catalogue. SOSIG also allows users to search on resources that are located in distinct geographic areas

Additional Subject-Related Services: Some subject gateways have built a considerable number of related services and information sources around their core searchable/browsable gateway. Biz/ed, the Business and Economics information gateway for students, teachers and lecturers, contains large amounts of business and economic information, such as company financial data and economic datasets from sources such as the office of National Statistics.

Resource Cataloguing: The key difference between subject gateways and the popular automated large-scale web indexing systems such as AltaVista is the quality of the results, which the end-user receives.

Somme important subject gateways in the area of Social Sciences and Humanities are:

Social Science Information Gateway (SOSIG)

(<http://sosig.esrc.bris.ac.uk>) - The Social Science Information Gateway (SOSIG) the best information gateway in the area of social sciences. It is a freely available Internet service which aims to provide a trusted source of selected, high quality Internet information for students, academics, researchers and practitioners in the social sciences, business and law. It is part of the UK Resource Discovery Network. SOSIG is a kind of one stop for navigators looking for social science resources.

Humbul (<http://www.humbul.ac.uk/>)

An excellent source in the humanities, covering large collection of high quality links to scholarly resources in the humanities. The Humbul Humanities Hub is a service of the Resource Discovery Network funded by the Joint Information Systems Committee and the Arts and Humanities Research Board, and is hosted by the University of Oxford. The Humbul Humanities Hub aims to be UK higher and further education's first choice for accessing online humanities resources. It provides pointers to resources on a wide dimension of areas in humanities. Besides, it includes details about the annual Digital Resources for the Humanities (DRH) conference, which is the major forum for all those involved in, and affected by, the digitization of our cultural heritage. It also provides quick links to resources for humanities research pertaining to various sub-fields.

- 9) On-line Writing Guides/Labs resources are very valuable for the Humanities areas as ready reference tools that provide tips and/or advice for writing various types of essays, constructing an argument, and the like. Examples include the following:

(http://www.pol.adfa.oz.au/resources/essay_writing/contents.html)

A thorough, indexed handbook to writing essays with helpful explanations and advice. Resource references do not contain hyperlinks, which means you need to note the site addresses and use those addresses for using the document online.

Search Tools/ Facilities	Advantages	Disadvantages
Search Engines	Comprehensive coverage	Generally lack explanatory information about the retrieved materials
Meta-Search Engines	Useful where comprehensiveness is the aim	Generate huge number of hits. Hence, it takes time to sift through the search results
Subject Catalogues/ Directories	Provide better results than search engines, because of the human interface.	Not as comprehensive as search engines, owing to the level of human input required in selecting, evaluating and describing resources.
Rating & Reviewing Services	Use star ratings and numerical scores to indicate the quality of a site at a glance	Though many of these services form part of search engines and automatically expand a search using a search engine when no hits are generated using these services, however, the rating services do not appear as default option when a query is made to search engine.
Subject-based Gateway Service and Virtual Libraries	Provide access to detailed descriptions of high quality resources	Cover relatively small number of materials due to high level human involvement. Not updated as frequently as search engines

- 11) It is important for academicians and researchers undertaking research to keep up to date with the latest literature in their field. The wealth of information available to us is enormous. New resources are being added to Internet every now and then. Therefore, looking for a specific piece of information on the Internet is like searching for a needle in a haystack. How does one know which is the best site from where one can get information that one is looking for. One needs sources to help make this task easier. Search tools, resource directories and current awareness services offer some solution. It is important to keep track of new sites that are added in the concerned area of specialization. Keeping up-to-date is essentially a matter of following appropriate discussion lists and Usenet newsgroups. Other people recommend or announce new resources, which they find on the Net or have used. Similarly, you can bookmark or preserve URL of a source in a file for future use, which you think, is appropriate. However, there are some sources available on the Internet, which keep you posted about new resources. Examples include: Internet Resources Newsletter (<http://www.hw.ac.uk/libWWW/irm/aboutirm.html>); Scout

Report (<http://re/internic.net/scout/report>); WWW Social Sciences Newsletter (<http://www.clas.ufl.edu/users/gthursby/socsci/news.htm>; BUBL LINK Updates

(<http://bubl.ac.uk/link/updates/current.html>); HUMBUL: new resources (<http://www.humbul.ac.uk/output/new.php?int=7>);

Infomine: scholarly Internet resource collections (<http://infomine.ucr.edu/>).

15.8 KEYWORDS

- | | |
|---|---|
| Browsers | : A program, which used to view information available via networks. Example are Netscape and Internet Explorer Which are browsers used for examining information available via the WWW. |
| Bookmark | : A website address stored on a user's computer as a menu item so that he or she can easily return to it without having to retype the address, referred to as a 'favorite' in Internet Explorer. |
| Campus Wide Information Systems (CWIS) | : Campus-Wide Information System is an information system intended to present an integrated view of the institution to the members of its community, as well as to alumnae, prospective students and others with an interest. Generally, within the menu hierarchy, a broad array of local and Internet resources are made available. The CWIS is a growing resource always under construction. |
| Courseware Directories | : There is a wide variety of courseware resources available via the Internet. These are very valuable resources. |
| Data Archives | : A data archive is a permanent, electronic collection of datasets with accompanying metadata such that users of the data can acquire, understand, and use the data. |
| Database | : A collection of records, each with details of a different data items, whether numeric, textual or image-based; usually searchable. |
| Discussion Lists | : Also some times called a 'mailing list' or 'listserv'. A discussion list is circulated to a group of e-mail uses who are interested in a particular topic/ area; e-mail users subscribe to the list, and can then post to it and receive all the message which are posted. |
| Electronic Journals | : A journal which is produced n an electronic format; sometimes the electronic equivalent of a paper-based journal, although an increasing number of journals are produced entirely in an electronic format. |
| E-Zines | : Are electronic magazines. |
| Gateway Service | : A service allowing authorised material such as e-mail to pass between dissimilar networks, by dovetailing one network's protocol with the others. |
| Hypertext | : A method of organising information retrieval that brings together related material. Hypertext lets a user access words and sections in various, scattered documents or files without |

having to shut the file the user is in. For example, World Wide Web, as a hypertext program, connects all the bits and pieces of related information from files all over the world and brings them together in a web of links.

Invisible Web

- : Sometimes also referred to as the 'deep web' the information that cannot be accessed by search engines because it is hidden behind a log-in screen or a search interface.

Mailing List or

Listserv

- : Also sometimes called a 'discussion list' or 'listserv', a mailing list is a group of e-mail users who are interested in a particular topic area; e-mail users subscribe to the list and they can then post to it and receive all the messages which are posted.

Meta Resources

- : Meta resource is a database of Internet resource descriptions that is made accessible through a structured and/or unstructured network service.

Meta-search Engines

- : Also sometimes called 'mete-crawlers' and 'multi-search engines', these allow users to search several search facilities once at the same time.

Multimedia

- : The ultimate computing experience, a combination of two or more media— such as text, still pictures, audio, video - presented in an interactive setting.

Navigators

- : Finding your way around in cyberspace.

Netscape Navigator

- : A multimedia browser for accessing information which is available via the WWW.

Preprints

- : Preprints most often refer to manuscripts that have not yet been published, but may have been reviewed and accepted; submitted for publication; or intended for publication and being circulated for comment. A preprint accessible over the web may also be referred to as an 'e-print'.

Resource Guides

- : The JISC is establishing posts known as Resource Guides in a number of subject areas. These posts will take over the maintenance of lists of resources known as Subject Guides.

Subject portals

- : Portals pertaining to specific subjects.

Subject-based

- : A collection of searchable (and usually browsable) resources

Gateway Services

- descriptions which have been developed by library professionals and/or subject experts with the explicit aim of providing access to high-quality sources of information. The descriptions relate only to the high-quality materials available via the Internet within a particular subject area.

Technical Reports

- : A report, which provides technical data collated by a committee, which contains material, not considered suitable or appropriate for a standard.

Uniform Resource

Locator (URL)

- : The full address of a resource on the Internet; every site or page on the Internet has a unique URL which consists of the protocol (e.g. http), the server name, the domain name, and the pathname of the resource.

: The part of the Internet based around HTML documents; hypertext links enable users to move between different parts of the same documents, as well as to different documents in different locations.

15.9 REFERENCES AND FURTHER READING

Berson, Michael J. (ed), [et al.]. (2000). *Social Studies on the Internet*. Illinois: Waveland Press.

Broadbent, Brooke (2002). *ABCs of E-Learning: Reaping the Benefits and Avoiding the Pitfalls*. USA: Jossey-Bass.

Cohen, Barbara (1997). *Social Studies Resources on the Internet: A Guide for Teachers*. UK: Heinemann.

Cooke, A. (2001). *A Guide to Finding Quality Information on the Internet: Selection and Evaluation Strategies*. 2d ed. London: Library Association Publishing.

Doherty, Paul (2000). *Cyberplaces: The Internet Guide for A/E/C*. Kington HA, USA: Robert S. Means Co.

Finding Information on the Internet: A Tutorial. (<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html>)

Gilbert, Jill S, [et al]. (2001). *Online: Investing Bible*. New York: John Wiley & Sons.

Gordon, Rachel Singer (2001). *Teaching the Internet in Libraries*. Chicago: American Library Association.

Jones, D. (1999). *Critical Thinking in an Online World*. Available at : <http://www.library.ucsb.edu/untangle/jones.html>

Joseph, Linda C.(1999). *Net Curriculum: An Educator's Guide to Using the Internet*. Medford, NJ, USA: Information Today Inc.

Kovacs, D.K. [et al.](1994). A Model for Planning and Providing Reference Services Using Internet Resources. *Library Trends*. 42(4), 638-47.

Searching for Information on the Internet. Available at: <http://www.netskills.ac.uk/support/Searching/search1.html>

Sherman, Chris and Price, Gary (2001). *The Invisible Web: Uncovering Information Sources Search Engines Can't See*. Medford, N.J., USA: Cyber Age Books.

Tseng, G., [et.al.]. (1997). *The Library and Information Professional's Guide to the Internet*. 2nd ed. London: Library Association Publishing.

UNIT 16 EVALUATION OF INTERNET RESOURCES

Structure

- 16.0 Objectives
- 16.1 Introduction
- 16.2 Need for Evaluation
- 16.3 Quality Assessment
- 16.4 Evaluation Tools on the Net
- 16.5 Evaluating Information Resources
 - 16.5.1 Types of Evaluation Criteria
 - 16.5.2 Generic Criteria for Evaluation
 - 16.5.3 Specific Criteria for Evaluation
- 16.6 Summary
- 16.7 Answers to Self Check Exercises
- 16.8 Keywords
- 16.9 References and Further Reading

16.0 OBJECTIVES

With so much of information being added to the web, it becomes imperative to sieve Internet information resources before disseminating to the end user. Libraries and information centres need to draw upon their expertise to identify, select, evaluate and sift Internet resources so as to provide their clientele access to only quality resources. After going through this Unit you will be able to:

- understand the need for evaluation of Internet resources;
- know various factors for assessment of quality;
- learn how to determine the relevance and authenticity of a given resource;
- discover how to find quality resources based on the application of some evaluation criteria; and
- find out how to critically analyse information sources.

16.1 INTRODUCTION

Information exists on a continuum of reliability and quality. The web has become the largest available repository of multimedia information. There is a vastly wide variety of material on the Internet, ranging in its accuracy, reliability, and value. The huge growth and the fast rate of change of the web makes it really hard to retrieve information relevant to specific user's needs. Web data are typically heterogeneous, unlabelled, time varying and distributed. Hence, identifying quality content, identifying and removing malevolent content and linking and subsequently user feedback are important components to be addressed here. Unlike most traditional information media such as books, magazines, organisational documents, no one has to approve the content of Internet before it is made public. It's your job as a searcher, then, to evaluate what you locate, in order to determine whether it suits your needs.

Just like you begin evaluating a physical information source – a book or an article for instance – even before you have the physical item in hand by first examining the bibliographic citation, which helps you to determine the usefulness of the source for your paper or work. In the same way, you can evaluate a website by examining the home page carefully.

16.2 NEED FOR EVALUATION

As stated earlier, there is colossal information available on the web. The World Wide Web can be a great place to accomplish research on many topics. But, putting documents or pages on the web is easy, cheap or free, unregulated, and unmonitored. The great wealth that the Internet has brought to so much of society is the ability for people to express themselves, find one another, exchange ideas, discover possible peers worldwide they never would have otherwise met, and, through hypertext links in web pages, suggest so many other people's ideas and personalities to anyone who comes and clicks. Therein lies the rationale for evaluating carefully whatever you find on the Web. The burden is on you as the information provider, on the reader so as to establish the validity, authorship, timeliness, and integrity of what you find. Documents can easily be copied and falsified or copied with omissions and errors, which may be intentional or accidental. Like most print publications, in the general World Wide Web there are no editors to proofread and send it back for updation/corrections or reject the stuff until it meets the standards of a publishing house's reputation. Most pages found in general search engines for the web are self-published or published by businesses small and large with motives to get you to buy something or believe a point of view. Even within university and library websites, there can be many pages that the institution does not try to oversee. This is by and large expected, as the web needs to be free like that. But if you want to use it for serious research, you need to cultivate the habit of healthy skepticism, of questioning everything you find with critical thinking.

The resources chosen and made accessible are to serve the specific research and teaching interests of the institution's user community.

Hence, broadly speaking it is important to subject Internet resources to an evaluation process so as to:

- Verify authenticity of the resources;
- Identify the authority of the author who is responsible for the content. It may just be possible that person whom the credit goes for the content is not an authority in the given subject area, which to a certain extent helps in deciphering the quality of the content. On the other hand, an accredited author that could be an individual or more appropriately an organisation would ensure quality and stability of the information;
- Find out validity of information in terms of being recent or outdated; and
- Decipher the relevance of the given resource for the intended user group.

16.3 QUALITY ASSESSMENT

As assessment of quality is primarily dependent upon two core factors, namely,

- 1) the needs of the individual, and
- 2) nature of the source being evaluated.

Besides, extensive usage of a particular resource will also give you a fair indication about quality of the material. However in general, to know the quality of a source one should look for the following aspects:

- Currency and maintenance of the resource – how frequently the resource is updated;
- Scope and purpose of the source - what is the coverage and intended readership or audience of the source;

- Creator/producer of the source – whether it is produced by a reputed organisation/individual/NGO and the like; and
- Presentation, ease of usage and help in using related resources.

16.4 EVALUATION TOOLS ON THE NET

From a librarian standpoint, we need to use the same critical evaluative skills in looking for information on the Internet that we would do for a book, a paper, index, a musical score, or on an online commercial database. The content of the Internet is only more diverse because of the potential of interaction with more media.

With the growth of information on the Internet and the development of more sophisticated searching tools, there is now more possibility of finding information and answers to real questions. But, within the morass of networked data are both valuable nuggets and an incredible amount of junk.

How should users today approach searching on the net and critically evaluating the data they find? You need a systematic approach to evaluating the tools you will use for searching and what they will cause you to receive or keep you from receiving. At the same time you also need a systematic approach to evaluating the document or result that you receive as a result of your search. As information professionals we are in the best position to determine and expand the relevance of existing criteria to new and future formats.

Learning how to determine the relevance and authority of a given resource for your research or for providing relevant information to users is one of the core skills of the research process. There are a number of sites that direct and assist you in evaluating quality of Internet resources, such as the following:

Internet Detective (<http://sosig.ac.uk/desire/internet-detective.html>)

This is a very useful and entertaining inter-active tutorial for evaluating the quality of Internet resources.

Searchability: Guides to Specialised Search Engines (<http://www.searchability.com>)

A good descriptive guide to specialised search engines, and to search directories of search engines, evaluating their subject coverage and effectiveness.

Search IQ (<http://www.searchiq.com>)

Provides access to a very large number of search engines and directories, and combines this with independent search-engine reviews and rankings, to help you identify the right search tool for the job - and those with the highest IQ! Also offers tips to improve your searching techniques, with tutorials and guides.

Others

i) Intermediary Tools

Bschool.com, (<http://www.bschool.com/>): Originally, the Marr-Kirkwood business school rating guide. A review guide that describes the resource and puts its stamp of approval by the number of stars.

Yahoo! (<http://www.yahoo.com>) is one of the tools that helps you identify like resources so that you can compare them.

ii) Popular Search Engines

In a number of cases it makes more sense to search a popular search engine to go directly to material on your specific term than it does to browse through directories or review tools.

With these tools you are searching using the value of description rather than that of evaluation. Most search engines today have some sort of associated portal. Danny Sullivan's Search engine Watch (<http://www.searchenginewatch.com/>) and Greg Notess' Search Engine Showdown (<http://www.notess.com/search>) are two current tools for keeping abreast of search engine developments. There is no good advice as to which one search engine is the best. They keep changing constantly.

iii) Search Engine Partners

Search engine partners such as Ask Jeeves (<http://www.ask.com/>) (question answering resource) Teoma (<http://www.teoma.com/>) (subject specific popularity). Each adds a value depending upon its niche.

iv) Directory Partners

Check to see if your search engine has licensed one of these or is creating its own. Inktomi (<http://www.inktomi.com/>) - purchased by Yahoo in 2003 ; Look Smart. (<http://www.looksmart.com/>)

Sites are springing up that purport to provide 'evaluations' of Internet resources. The next thing that is needed is to evaluate their track record to determine the value of their evaluations. While there are criteria in each case, the implementation of the evaluations is frequently subjective or biased. It may be worth mentioning here that this is really no different than what we have lived with in the print environment, except that now it is digital.

v) General Guides and Directories

Such as WWW Virtual Library (<http://vlib.org/>); Yahoo! - (<http://www.yahoo.com/>)

Among the general guides there are a number of sites that purport to be the site you should start with: -

About.Com (<http://www.yahoo.com/>);

Webfile (<http://www.thebighub.com/>);

Search Engine Colossus (<http://www.searchenginecolossus.com/>)

You might want to compare in terms of value, to you the level of specificity in Yahoo and the WWW Virtual Library and the newer general directories versus the set of categories in the various directories of the search engines.

vi) Specialised Guides: Examples include:

A Business Researcher's Interests (<http://www.brint.com/>);

StreetEYE Supersites (<http://www.streeteye.com/index/index.html>);

Bschool.com (<http://www.bschoool.com/>);

OneLook Dictionary (<http://www.onelook.com/>)

vii) More traditional library resources (fee-based and generally worth the cost):

- Examples include: - EBSCOhost (<http://www.ebscohost.com/>);

First Search (<http://firstsearch.oclc.com/>);

Proquest (<http://www.proquest.com/>);

Thomson Gale (<http://www.gale.com/>)

Self Check Exercise

- 1) What are various tools available on the Net that can direct and assist you in evaluation process?
- 2) Where does the need for evaluation of information resources on the Net lie?

Note : i) Write your answers in the space given below.

- ii) Check your answers with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

16.5 EVALUATING INFORMATION RESOURCES

A key problem with the web is that it is carelessly used as a primary source of information, when in fact it's more often a secondary source. Someone reads an inaccurate statement on the Web, quotes it on another web page, search engines obligingly index all of the inaccurate pages, and we end up with a mess where fiction is accepted as reality. But perhaps one cannot take such things very lightly if you are using such resources for quality work.

Remember that the web is a public place, open to all. Anyone can publish on the Internet, so, information has often not been put through traditional quality 'filters' such as publishers, editors or peer-reviewers. Resources may not be what they appear to be or what they say they are, as on the Internet there is no guarantee that the resource is accurate or honest. You need to be aware of the entire range of human possibilities of intentions behind web pages.

Hence, while you are in for new resources they might seem to be useful and might fall within the ambit and scope of your requirements and objectives, such resources need then be evaluated to ascertain the actual worth (without populating your resource pool with alien resources) in terms of content, form and process. These criteria are not as easy to define as the scope criteria, and you will need to weigh up the strengths and weaknesses of a resource to form an overall judgment

16.5.1 Types of Evaluation Criteria

In the research process you would usually encounter many types of resources including books, articles and websites. But not everything you find on your topic is suitable. How do you evaluate its authority and appropriateness for your research? It is necessary to understand the current state of the Internet to determine how you best identify the quality of an Internet resource in this volatile, continually changing environment. .

We shall attempt here to broadly classify the evaluation criteria into two types based on the nature and dimension of the evaluation required. For instance, if you want to assess the quality of a resource to be utilised immediately, you may judge by looking at few indicators. However, such evaluation process should be treated purely as context-oriented.

The two types of evaluation criteria are:

- a) Generic Criteria
- b) Specific Criteria

It may be noted that under these two types of evaluation criteria there is bound to be some overlapping. We shall discuss now various aspects associated with each of these two types of criteria.

The schematic representation presents in Figures 16.1 and 16.2 some criteria for the evaluation.



Fig. 16.1: Generic Criteria for Evaluation

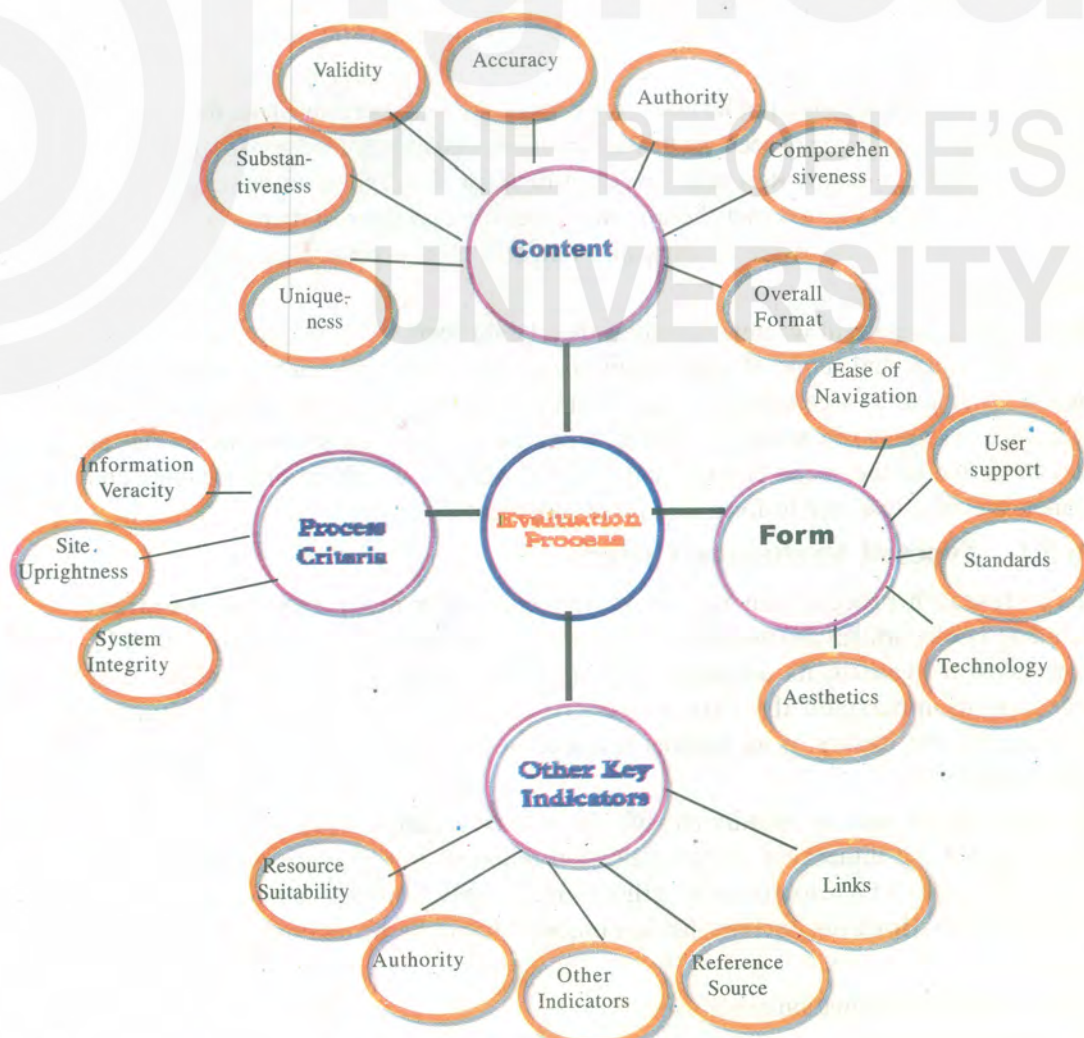


Fig. 16.2: Specific Criteria for Evaluation

16.5.2 Generic Criteria for Evaluation

To evaluate an Internet resource, you need to think critically, even suspiciously, by asking a series of questions that will help you decide how much of the resources displayed by a website are to be trusted.

The techniques that you will be employing for the process begins with looking at your search results from a search engine or other source, followed by investigating the content of page, and extends beyond the page to what others may say about the page or its author(s) credentials and other quality factors. We shall briefly discuss each one of these with a set of questions that come to your mind

a) URL Appropriateness

Before you leave the list of search results, glean all you can from the URLs of each page. Then choose the pages most likely to be reliable and authentic.

What can the URL tell you? It can help you in several ways such as whether the URL is appropriate. You can check:

- Is it somebody's *personal page*?
- Is the server a commercial ISP or other provider mostly of web page hosting like aol.com What type of domain does it come from ?
- URLs can also reveal bias, such as microsoftsucks.com.
- Who 'published' the page? Is it published by an entity that makes sense?

Personal pages are not necessarily bad, but you need to verify the author's (credentials) very carefully. For personal pages, there is no publisher or domain owner vouching for the information in the page. In general, perhaps a site hosted on a free server might not have the same credibility as one from a .gov or .edu site. Use Whois Domain Lookup (<http://www.networksolutions.com/cgi-bin/whois/whois>) to know who owns the domain. In general, the publisher is the agency or person operating the server computer from which the document is issued. The server is usually named in first portion of the URL (between http:// and the first /). For instance, you can coolly look for health information from any of the agencies of the National Institute of Health on sites with nih somewhere in the domain name.

b) Responsibility/ Author

The first thing that you should look for on this aspect, is:

- Who is responsible for the work to decipher authority or author c.: creator
- What are the author's credentials on this subject?
- Does the purported background or education look like one who is qualified to write on this topic?

Look for the name of the author, or the name of the organisation, institution, agency, or whatever is responsible for the page. An e-mail contact is not enough. If there is no personal author, look for an agency or organisation that claims responsibility for the page. An e-mail address with no additional information about the author is not sufficient for assessing the author's credentials. If you cannot find this, locate the publisher by truncating back the URL. You should hold the author to the same degree of credentials, authority, and documentation that you would expect from something published in a reputable print resource such as a book, journal article, good newspaper.

There should be an 'about us' page, with contact information and credentials. Any site that omits this information is of questionable authority.

c) **Purpose**

- Why was the page put on the web? (to: inform, give facts, give data, explain, persuade, sell, entice, share, disclose?)
- What is the tone of the page (ironic, satire, and the like)
- Is the page merely an opinion or it has some quality basis (from research point of view)?
- Is the page a rant, an extreme view, possibly distorted or exaggerated?

These are some of the reasons to think of. The web is a public place, open to all. You need to be aware of the entire range of human possibilities of intentions behind web pages. Anyone can put anything on the web for pennies in just a few minutes. Your task is to distinguish between the reliable and questionable. Many web pages are opinion pieces offered in a vast public forum.

d) **Documentation**

If you cannot find strong, relevant credentials, look very closely at documentation of sources.

- Are sources documented with footnotes or links?
- Who do they link to?
- Who links to them?
- Do the links work?
- Does it include copyright information?

In scholarly/research work, the credibility of most writings is proven through footnote documentation or other means of revealing the sources of information. Saying what the author believes without supporting documentation is not much better than just expressing an opinion or a point of view.

Check the external links to see what kind of places they lead to. Are some of them questionable? Do an AltaVista field search (link: "whatever the URL is") to find out who links to them. Links that don't work or are to other weak or fringe pages do not help strengthen the credibility of the page. Look at the bottom of such articles for copyright information or permissions to reproduce. If you find a legitimate article from a reputable journal or other publication, it should be accompanied by the copyright statement and/or permission to reprint. If it is not, be suspicious.

e) **Recognition by Others**

Try to determine whether the site has been recognised as exemplary, either in reviews or by others linking to it.

- Who links to the page?
- Is the page listed in one or more reputable directories or pages?
- What do others say about the author or responsible authoring body?

Find out how many links are there, what kinds of sites link to it, what do they say, are any of them directories and the like. Sometimes a page is linked to only by other parts of its own site (not much of a recommendation). Sometimes a page is linked to by its fan club, and by detractors. Read both points of view. If a page or its site is in a bonafide directory, think about whether there is much critical evaluation of the links in the directory.

Good directories include a tiny fraction of the web, and inclusion in a directory is therefore noteworthy. If the viewpoint about author or creator is radical or controversial, expect to

find detractors. Think critically about all points of view. Figure out the comparability with related sources from the links that it includes or is being linked to.

f) **Information**

The information component needs to be viewed from three angles:

- Content
- Currency
- Coverage

In terms of the content, one needs to ascertain the stability of information

Can I rely on it staying there?

Coverage – should focus on the scope and objectives of the content vis-à-vis the user requirements.

Currency – should generally be viewed in terms of the following:

- Is the page dated?
- Is it current enough?
- When was the last update?

Look for the date last updated - usually given at the bottom of a web page. Check the date on all the pages on the site. Do not rely on a date given in Internet Explorer's File Properties or Netscape's View Page Info displays. These dates can be automatically kept current and are useless in critical evaluation.

However, the requirement of currency depends on your needs. For some topics you want current information, while for others, you may want information put on the web near the time it became known.

g) **Personal Discretion**

Personal discretion would be based on common sense. Just because a major search engine may suggest the site, it doesn't mean it's accurate. Use the same kind of 'internal filter' you use when reading print material like newspaper and magazine articles and junk mail.

- Ask yourself if the web is truly the best place to find resources for the research you are doing.
- Is this as good as resources that I could find if I used the library, or some of the web-based indexes available through the library, or other print resources?
- What is your requirement or your user's requirement for the quality of reliability of your information?

16.5.3 **Specific Criteria for Evaluation**

Here, we shall be taking up the detailed and more specific aspects of evaluation. The various evaluation parameters or criteria have been broadly categorised under four categories as shown below:

a) **Content Criteria**

- Validity
- Authority
- Substantiveness

- Accuracy
 - Comprehensiveness
 - Uniqueness
 - Overall Format
- b) Form Criteria
- Ease of navigation
 - User support
 - Standards
 - Technology
 - Aesthetics
- c) Process Criteria
- Information veracity
 - Site uprightness
 - System integrity
- d) Other Key Indicators / Criteria
- Resource suitability
 - Authority
 - Reference sources
 - Links
 - Other indicators

a) **Content Criteria**

The information content is a primary consideration when evaluating Internet resources. The content of resources can be evaluated by considering the following criteria:

- i) **Validity** – Consider the following under this parameter / criteria
- Do the resources fulfill the stated purpose? (Find out the scope and missing information, if any)
 - Does the information appear to be well researched? (Whether sources of information such as references and bibliography are included?)
 - Is there any mention of the resource being available in another format?
 - What was the incentive of the information provider for making the information available? (Just for vanity, or other purposes)
 - Is the information genuine at least by what it appears to be?
Does the URL support the claim of authorship?
Does the resource point to other sources, which could be contacted for confirmation?

ii) **Authority**

Here, mainly authority and reputation of the source will be deciphered. Whether the resource is well documented or not. Where is it hosted? Who and how reputable is the information provider? (Credentials of the author/creator in terms of his reputation or is he frequently cited author, subject expert? etc.)

iii) **Substantiveness**

Does the resource contains more than contact details? Is the information full-text? Is it merely advertising? Do we get some value added information such as an annotated bibliography or things like that?

iv) **Accuracy**

Is the information accurate with regard to content, grammar and typographical errors.

v) **Comprehensiveness**

The extent of details the resource has is another important parameter of evaluation. The things one can look for this is whether the title is informative; an abstract is given; some sort of opening mission statement of the purpose of the resource is given; or keywords that indicate the information content are included. Besides, you need to see whether everything you expect to find in the site is there.

vi) **Uniqueness**

This parameter or criterion addresses the uniqueness or otherwise of information available on the site, This can be broadly judged by the nature of the material whether it is primary material; or any original work available at the site; or material having any relation to other works or whether is site inward focused which means it is not just a list of links to external sites but much more than that.

vii) **Overall Format**

Yet another important aspect is how well the information is composed and organised. Good composition with respect to following of basic rules of grammar, spelling and literary composition for the text and related aspects. Good organisation of information refers to good structure; logical and consistent arrangement of information within a resource that is user friendly; segmentation of information for easy consumption and registering; clarity in the content description; well framed meaningful headings, etc.

b) **Form Criteria**

This category of evaluation parameters or criteria is primarily user-specific concerned mainly with evaluation of the medium. Therefore, form criteria should be applied by looking at and evaluating the medium, design and presentation of the resource. Different people see Internet resources in different forms due to heterogeneous systems. As such, these criteria focus on user-friendly aspects that shall facilitate the user's navigational activity and on ease of use. The form of resources can be evaluated by considering the following criteria:

i) **Ease of Navigation**

This refers as to how easy it is to navigate the resource. Whether the links, graphics are clearly identifiable and easily navigable; do print options have convenient layout, and so on. For instance, whether index is available or can be generated automatically. Besides, how effective and efficient search facility is provided.

ii) **User Support**

Does the site provide for adequate essential instructions; online help; online documentation and other requisite online customer support?

iii) **Standards**

Whether use of recognised standards such as those of metadata have been used: multimedia formats like MIME used; standard language such as HTML with requisite extensions that all browsers recognise has been used, and so on.

iv) Technology

Appropriate use of technology is another criterion in finding quality of information. Here, one could observe whether appropriate interactivity is available.

v) Aesthetics

As the term indicates, here one could view as to whether due consideration has been given to the appearance of the site. Good design principles have been followed; look and feel of the site is aesthetically sound; sound balance between colour schemes used, graphics, links and the like.

Self Check Exercise

- 3) Name the two broad types of evaluation criteria. What are the aspects associated with the Generic Criteria of evaluation?

- Note :** i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

c) Process Criteria

These criteria are mainly concerned with aspects of system evaluation. The resources on the Internet are volatile and likely to change over time and the integrity of Internet resources is dependent on three driving forces - original information provider, the website manager, and the underlying technology. The process criteria need to be applied by considering the processes and systems, which exist to support the information resource. The information, the interface and the system supporting the resource should all be reliable. The processes associated with resources can be evaluated by considering the following criteria:

i) Information Veracity

Information veracity is basically the work of the information provider. Here, the relevant evaluative questions can be:

- Is the information current and up to date?
- How durable is the nature of the information?
- Is there adequate maintenance of the information content? and so on.

ii) Site Uprightness

The work of the website manager should address issues like regular updation of the site, removing all dead links, referring to corresponding links if an existing URL is dysfunctional and things like that. He should effectively administer and maintain the site so that durability and ongoing maintenance and stability of the resource is ensured.

iii) System Integrity

System integrity being the work of the system administrator, he should focus on issues as to whether technical performance of the resource is acceptable; stability of the system is ensured and necessary measures to maintain the integrity of the system are taken, for instance whether the site is mirrored, and the like.

d) Other Key Indicators

The other key indicators that are crucial for critical evaluation of resources include the following: -

i) Resource Suitability

Suitability in terms of -

Scope - whether the resource covers the right stuff in right proportion that the user is interested in.

Audience - Intended audience for this source. For instance, to find resources listing the latest research trends in social sciences of interest to medical professionals, you may want to avoid the SOSIG (Social Sciences Gateway on the Net) which will bring up information on this for social science researchers rather than for medical practitioners.

Timeliness - Avoid using undated websites. Always check the date when source was published or the date last updated if it is a website. Library catalogues and periodical indexes always indicate the publication date in the bibliographic citation.

Scholarly versus Popular

A scholarly journal is generally one that is published by and for experts and all articles are peer-reviewed that ensures quality of resource. Scholarly sources would usually include: bibliography and footnotes; author's name and academic credentials.

Popular magazines range from highly respected publications such as Scientific American and The Atlantic Monthly, to general interest newsmagazines like Newsweek and US News and World Report. Articles in these publications are usually written by staff writers or freelance journalists and are geared towards general audience. While most magazines adhere to editorial standards, articles do not go through a peer-review process and rarely contain bibliographic citations.

Try to conduct your searches in journal indexes and try narrowing your search by limiting to refereed publications. This will retrieve only scholarly journals matching your search terms. If you do your searches in Web of Science, Current Contents, you will retrieve only scholarly articles since only academic journals are indexed in these databases.

ii) Authority

The author (s) academic credentials are crucial in deciphering the worth of the resource. This aspect has already been discussed earlier. There are reference sources available that list biographical information about authors and scholars.

iii) Other Indicators

Other indicators for quality evaluation include the following: -

Documentation - A bibliography, along with footnotes, indicate that the author has consulted other sources and it serves to authenticate the information that he or she is presenting. In websites, expect links or footnotes documenting sources, and referring to additional resources and other viewpoints.

Objectivity - Information content of a resource should be factual, unbiased and written objectively. It should not be a propaganda, non-encouraging advertisements in the website sponsored by a company that advocates a certain philosophy. There are resources such as magazines for libraries, etc. that can be consulted which indicate whether a publication is known to be conservative or progressive, or is affiliated with a particular advocacy group.

Primary Versus Secondary Research — In determining the appropriateness of a resource, it may be helpful to determine whether it is primary research or secondary research. Primary research presents original research methods or findings for the first time. Examples include: journal article, report, or other publication or a newspaper account written by a journalist

A secondary research does not present new research but rather provides a compilation or evaluation of previously presented material. Examples include: an encyclopedia entry and entries in most other reference books and textbooks.

Self Check Exercise

- 4) What are the criteria used to judge the information resource suitability?

Note : i) Write your answer in the space given below.

- ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

iv) Reference Sources

Many reference sources provide evaluative description about various sources such as the following:

Book Reviews A book review can appear in a journal, magazine or newspaper, which provides a descriptive, evaluative discussion of a recently published book. Reading how others have evaluated a book may help you decide whether to use that book in your research or not. There are a number of indexes you can consult that provide references to book reviews: Book Review Digest (1905-current) available also Online for 1983-current; Book Review Index (1965-)

Journal Indexes You may also find reviews of books in many journal indexes in a given subject area by searching on the title and/or author of the book.

Citation Indexes To see the impact a particular source has had on scholarship, you may want to consult a citation index. A citation index lists when and where a work has been cited. The citation indexes are all available in the following database: Web of Science: <http://www.webofscience.com>; Science Citation Index (1945-present); Social Sciences Citation Index (1970-present); Arts & Humanities Citation Index (1975-present) The citation indexes are available in print, CD's and also online.

v) Links

The links to and from quality resources to the given site depicts the quality of the resource.

Self Check Exercise

- 5) How are process criteria important for the evaluation of a resource? What elements can be verified under using these criteria?

- Note :** i) Write your answer in the space given below.
ii) Check your answer with the answers given at the end of the Unit.

.....

.....

.....

.....

.....

.....

16.6 SUMMARY

Information exists on a continuum of reliability and quality. The web has become the largest available repository of multimedia information. There is an extremely wide variety of material on the Internet, ranging in its accuracy, reliability, and value. The huge growth and the fast rate of change of the web makes it really hard to retrieve information relevant to specific user's needs. Web data are typically heterogeneous, unlabelled, time varying and distributed. Hence identifying quality content, identifying and removing malevolent content and linking and subsequently user feedback are important components to be addressed here. Unlike most traditional information media such as books, magazines, organisational documents, no one has to approve the content of before it is made public. It's your job as a searcher, then, to evaluate what you locate, in order to determine whether it suits your needs. The World Wide Web can be a great place to accomplish research on many topics. But, putting documents or pages on the web is easy, cheap or free, unregulated, and unmonitored.

Therein lies the rationale for evaluating carefully whatever you find on the Web. The burden is on you as the information provider, on the reader so as to establish the validity, authorship, timeliness, and integrity of what you find. Documents can easily be copied and falsified or copied with omissions and errors, which may be intentional or accidental. Hence, it is important to subject Internet resources to an evaluation process so as to:

- Verify authenticity of the resources
- Identify the authority of the author who is responsible for the content.
- Find out validity of information in terms of its recency or outdatedness
- Decipher the relevance of the given resource for the intended user group.

The assessment of quality is primarily dependent upon two core factors:

- i) The needs of the individual, and
- ii) Nature of the source being evaluated.

Learning how to determine the relevance and authority of a given resource for your research or for providing relevant information to users is one of the core skills of the research process. There are number of sites that direct and assist you in evaluating quality of Internet resources, such as the following:

- **Internet Detective** (<http://sosig.ac.uk/desire/internet-detective.html>)
- **Searchability: Guides to Specialized Search Engines** (<http://www.searchability.com>)
- **Search IQ** (<http://www.searchiq.com>)

● Other Intermediary Tools

An attempt has been made to broadly classify the evaluation criteria types into two based on the nature and dimension of the evaluation required. For instance if you want to assess the quality of a resource to be utilized immediately, you may judge by looking at few indicators. However, such evaluation process should be treated purely as context-oriented.

The two evaluation process criteria are:

- a) Generic Criteria
- b) Specific Criteria

The techniques that you will be employing for the process begins with looking at your search results from a search engine or other source, followed through by investigating the content of page, and extends beyond the page to what others may say about the page or its author(s) credentials and other quality factors. The Unit also provides questions that an evaluator of the Internet information resources needs to consider during the process of evaluation while applying any evaluation criteria.

The Generic Criteria for Evaluation involves the following steps:

- i) URL Appropriateness
- ii) Responsibility/ Author
- iii) Purpose
- iv) Documentation
- v) Recognition by Others
- vi) Information
 - a) Content
 - b) Currency
 - c) Coverage
- vii) Personal Discretion

The Specific Criteria for Evaluation involve detailed and more specific aspects of evaluation. The various evaluation parameters / criteria have been broadly categorized under four categories as given below.

- a) Content Criteria
 - Validity
 - Authority
 - Substantiveness
 - Accuracy
 - Comprehensiveness
 - Uniqueness
 - Overall Format
- b) Form Criteria
 - Ease of Navigation
 - User Support
 - Standards
 - Technology
 - Aesthetics

- c) Process Criteria
 - Information Veracity
 - Site Uprightness
 - System Integrity
- d) Other Key Indicators
 - Resource Suitability
 - Authority
 - Reference Sources
 - Links
 - Other Indicators

16.7 ANSWERS TO SELF CHECK EXERCISES

- 1) There are number of sites that direct and assist you in evaluating quality of Internet resources, such as:
 - Internet Detective (<http://sosig.ac.uk/desire/internet-detective.html>)
 - Searchability: Guides to Specialised Search Engines (<http://www.searchability.com>)
 - Search IQ (<http://www.searchiq.com>)
 - Other Intermediary Tools
- 2) Publishing documents or pages on the web is easy, cheap or free, unregulated, and unmonitored. The great benefit and opportunity that the Internet has brought to so much of the society is the ability for people to express themselves, find one another, exchange ideas, discover possible peers worldwide they never would have otherwise met, and, through hypertext links in web pages, suggest so many other people's ideas and personalities to anyone who comes and clicks. Documents can easily be copied and falsified or copied with omissions and errors, which may be intentional or accidental. Like most print publications, in the general World Wide Web there are no editors to proofread and send it back for updation/corrections or reject the stuff until it meets the standards of a publishing house's reputation.

Therein lies the rationale for evaluating carefully whatever you find on the Web. The burden is on you as the information provider, on the reader so as to establish the validity, authorship, timeliness, and integrity of what you find.

Hence it is important to subject Internet resources to an evaluation process so as to:

 - Verify authenticity of the resources
 - Identify the authority of the author who is responsible for the content.
 - Find out validity of information in terms of being recent or outdated
 - Decipher the relevance of the given resource for the intended user group.
- 3) The two types of evaluation criteria are :
 - Generic Criteria
 - Specific Criteria

The various aspects associated with Generic Criteria for Evaluation are:

- a) URL Appropriateness
 - b) Responsibility/ Author
 - c) Purpose
 - d) Documentation
 - e) Recognition by others
 - f) Information
 - i) Content
 - ii) Currency
 - iii) Coverage
 - g) Personal Discretion
- 4) The criteria for judging the suitability of a resource are:
- a) **Scope** – Whether the resource covers the right stuff in right proportion that the user is interested in.
 - b) **Audience** – Who are the intended audience for this source.
 - c) **Timeliness** - How current or outdated the source is. Currency of the site i.e.; date when the source was published or date last updated if it is a website.
 - d) **Scholarly vs. Popular** – Whether the source is popular or scholarly one. A scholarly journal is generally one that is published by and for experts and all articles are peer-reviewed that ensures quality of the resource. In popular magazines, articles are usually written by staff writers or freelance journalists and are geared towards general audience.
- 5) Process Criteria are mainly concerned with aspects of System evaluation. The resources on the Internet are volatile and likely to change over time and the integrity of Internet resources is dependent on three driving forces - original information provider, the Website manager, and the underlying technology. The process criteria need to be applied by considering the processes and systems, which exist to support the information resource. The information, the interface and the system supporting the resource should all be reliable. The following elements can be verified under these criteria:
- i) **Information Veracity** - Information veracity is basically the work of the Information provider. The elements / aspects that can be verified using these criteria are:
 - Is the information current and up-to-date?
 - How durable is the nature of the information?
 - Is there adequate maintenance of the information content?
 - Other similar aspects.
 - ii) **Site Uprightness** - The work of the website manager should address issues like regular updation of the site, removing all dead links, referring to corresponding links if the existing URL is dysfunctional and things like that. He should effectively administer and maintain the site so that durability and ongoing maintenance and stability of the resource is ensured.

- iii) **System Integrity** - System integrity being the work of the System Administrator, he should focus on issues as to whether technical performance of the resource is acceptable; stability of the system is ensured and necessary measures to maintain the integrity of the system are taken, for instance, whether the site is mirrored, and the like.

16.8 KEYWORDS

- Website** : A collection of linked HTML pages available via the WWW generally owned or produced by a single institution or individual
- Citation Index** : An index covering published articles, giving the bibliographic details of subsequent articles which cite each of the original articles.

16.9 REFERENCES AND FURTHER READING

Beck, S. (1997). *Evaluation Criteria: The Good, the Bad and the Ugly: Or Why is it Good Idea to Evaluate Web Sources*. Available online (<http://lib.nmsu.edu/instruction/evalcrit.html>).

Cooke, A. (2001). *A Guide to Finding Quality Information on the Internet: Selection and Evaluation Strategies*, 2nd ed. London: Library Association Publishing.

Critical Evaluation of Resources. (2002). Available online (<http://www.lib.berkeley.edu/TeachingLib/Guides/Evaluation.html>).

Doherty, Paul. (2000). *Cyberplaces: The Internet Guide for A/E/C*. Kingston, H A., U.S.A.: Robert S. Means Co.

Evaluating Information Found on the Internet. (2005). Available online <http://www.library.jhu.edu/elp/useit/evaluate/>.

Evaluation of Internet Resources in the Librarians' Index to the Internet (2002). Available online (<http://lii.org>).

Jones, D. (1999). *Critical Thinking in an Online World*. Available online (<http://www.library.ucsb.edu/untangle/jones.html>).

Meyer, E. Barbara. (2001). *Evaluating Web Sites*. EdTech Center-Life Sciences University of Illinois - Urbana, IL USA. Available online (<http://www.uiuc.edu/>).



ignou
THE PEOPLE'S
UNIVERSITY