Indira Gandhi
National Open University
School of Social Sciences

BLI-224
ICT Fundamentals

Block

4

INTERNET TOOLS AND SERVICES

Programme Design Committee

Prof. Uma Kanjilal (Chairperson)
Faculty of LIS, SOSS, IGNOU

Prof. B.K.Sen, Retired Scientist, NISCAIR
New Delhi

Prof. K.S. Raghavan, DRTC
Indian Statistical Institute, Bangalore

Prof. Krishan Kumar, Retired Professor
Dept. of LIS, University of Delhi, Delhi

Prof. M.M. Kashyap, Retired Professor
Dept. of LIS, University of Delhi, Delhi

Prof. R.Satyanarayana
Retired Professor, Faculty of LIS, SOSS, IGNOU

Dr. R.Sevukan (Former Faculty Member)
Faculty of LIS, SOSS, IGNOU

Prof. S.B. Ghosh, Retired Professor
Faculty of LIS, SOSS, IGNOU

Prof. T. Viswanathan, Retired Director
NISCAIR, New Delhi

Dr. Zuchamo Yanthan
Faculty of LIS, SOSS, IGNOU

Conveners:

Dr. Jaideep Sharma
Faculty of LIS, SOSS, IGNOU

Prof. Neena Talwar Kanungo
Faculty of LIS, SOSS, IGNOU

Programme Coordinators

Prof. Jaideep Sharma and
Prof. Neena Talwar Kanungo

Course Coordinator

Prof. Uma Kanjilal

Programme Editor

Prof. Jaideep Sharma

Course Preparation Team

Unit No(s).	Contributor(s)
13-16	Dr. Aditya Tripathi

Course Editor
Prof. Uma Kanjilal

Internal Faculty:
Prof. Uma Kanjilal
Prof. Neena Talwar Kanungo

Material Production

Mr. Manjit Singh
Section Officer (Pub.)
SOSS, IGNOU

Secretarial Assistance

Ms. Sunita Soni
SOSS
IGNOU

Cover Design

Ms. Ruchi Sethi
Web Designer
E Gyankosh
IGNOU

July, 2014

ISBN : 978-81-266-6773-4

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068 or visit University's Website http://www.ignou.ac.in.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by Director, SOSS.

Lasertypesetted at Graphic Printers, 204, Pankaj Tower, Mayur Vihar, Phase-I, Delhi-110091.

Printed at :

BLOCK 4 INTERNET TOOLS AND SERVICES

Introduction

Internet has been around for quite some time now and has revolutionised computer and communication world. Its popularity has increased to a great extent in the recent years. Internet is a information and communication channel that enables broadcast and exchange of electronic information amongst individuals and institutions. The usage varies from a simple e-mail to a complex web conferencing system.

Internet provides several services viz. searching for information through search engines or directories, communication through email, chat etc., file transfer and media broadcast. World Wide Web is the most popular service of the Internet which majority of people around the world are using. A variety of tools have evolved over the years to make effective use of these services. In this block we will study different tools and services of the Internet.

This block has four Units:

Unit 13 on **Email and E-messaging** deals with email service and providers, protocols and web 2.0 tools related to email

Unit 14 covers **World Wide Web (WWW)**, its development, technology, Web 2.0 services and its impact.

Unit 15 deals with **Search Engines** covering different search tools, search engines and technology behind it and using search engines for effective information retrieval.

Unit 16 discusses **Interactive and Distributive Services** ranging from basic distributive services like web directory, bulletin board, web portal etc. to more interactive services such as e-learning, e-marketing, file transfer, online document delivery etc.

Blank Page

UNIT 13 E-MAIL AND E-MESSAGING

Structure

13.0 OBJECTIVES

After reading this Unit, you will be able to:

- provide a detailed account about Email and Email service Providers;
- explain in detail various Protocols used in Email service; and
- discuss about Web 2.0 tools in Email.

13.1 INTRODUCTION

Electronic Mail is one of the most prominent uses of networked communication technology. Better known as email, this is one of the most widely used forms of communication today. Electronic mail (email) has many advantages over other forms of communication: it is easy to use, free of charge, fast, and delivers information in a digital format. With suitable encoding methods, email can be used to send any kind of computer file, including pictures, sounds, programs, and movies.

Email started in 1965 as a way for multiple users of a time-sharing mainframe computer to communicate among themselves. ARPANET had a significant role in popularising email. The foundation for today's global Internet email service was created in the early ARPANET.

13.2 EMAIL

Email is the most popular and the quickest method of transferring message over network. There are several service providers over WWW, offering free email services. However, organisation can offer own mailing service for the stakeholders of organisation. The service is so popular and effective that many of the organisations operate only on email for internal and external communication.

13.2.1 Defining Email

Electronic mail, or email, is the transmission of text-based messages among networked computers. Email is one of the earliest and most basic messaging resources on the Internet and in many ways it still acts as the lowest common denominator for computer communications.

Features

- It is faster and more secure than conventional mail.
- It requires less physical effort to edit and send a letter of communication.
- Once the hardware, software, and Internet connection are in place, email on the Internet is free, even if messages is to be sent to the other side of the world.
- Unlike communication by telephone, email does not require the attention of both parties at the same time.

13.2.2 Needs of Email

Email has become an important and integral part for the people who are living away or who has someone far away. For different persons the reasons are different towards using email. In general the reasons are:

1) An email ensures faster/easier delivery of messages as long as email address is correct.

2) It provides time-stamped proof of an interaction. Also, many email services (such as Gmail) collate the conversation on the same subject into single threads.

3) It is more secure and inexpensive compared to other modes of communication.

4) It is easy to archive for future recall. Most of the email services provide search facility through emails.

5) An email can be edited and rephrased as much as it is desired before sending to the recipient(s).

6) It is easy to send the same piece of information to several people simultaneously such as circulation of memos, agendas, and minutes, or disseminate educational material.

13.2.3 Email Address

An email address is a unique address, which identifies a location to send and receive email. The email address contains ***username***, followed by an @ symbol, and then domain name i.e. username@domainname

e.g. abc@yahoo.co.in

An email address starts with a **user name** (*abc* in this case) that refers to the recipient's mailbox. Then, **sign** @ followed by the **host name** (*yahoo.co.in* in this case), also known as **domain name**. Normally, the Domain name has three parts (two part in case of United States) separated by two period (.) symbols. Reading from the left the domain name the first part is *yahoo*, is the name of a machine, which is a **mail server** or the computer where the recipient has an electronic mailbox. The first part of the **domain name** ends with period followed by rest of the part known as top-level domain (TLD). The TLD may have two parts the first part represents the type of organisation and the second part represents country code (according to the name of the country). In the given example, ***co.in*** is the top-level domain, where ***.co*** qualifies that ***yahoo*** is a company. This part of the domain name indicates the type of organisation (***.com*** represents commercial organisation whereas ***.gov*** refers to a governmental setup). The last part represents the country where ***yahoo*** has registered this machine or hosted. It is two characters long country code. In the given example, it is ***.in*** which represents that ***yahoo*** has registered/hosted this machine in India.

13.3 TYPES OF EMAIL SERVICES

There are different types of email services based on the type of provider organisation and the terms and conditions.

13.3.1 Free Web-based Email Services

These are the commonly used email accounts accessible through web browser (such Internet Explorer, Firefox etc.). They generally use HTTP Protocol for accessing mail. Example: Yahoo Mail, Hotmail, Gmail offer free web based email services. These services are run on financial profit from advertisements. Advertisers pay service providers to expose email account holders to advertisement.

The service providers give users a document to accept terms and conditions for use of service, which, one must before signing for the service. These services have a provision

for spam filtration and virus scanning. Almost all the major services provide secure login using HTTPS (Hypertext Transfer Protocol Secure). HTTPS can be identified in the URL line while signing into the account (https://gmail.com or https://mail.yahoo.com).

13.3.2 Priced Web-based Email Services

There are also priced web based email service providers. Many of the free web based email service providers such as Yahoo! And Hotmail do offer premium account on payment basis. Advantage of utilising such services include more secured transaction of communication; personalised email address; better spam filteration as well as increased storage space. Priced email services are mostly useful in business related message exchanges.

13.3.3 Private Email Services

Most of the institutions/organisations have their own dedicated mail server and offer mail account for free to their staff and other members of the institution. These accounts exist only till one is member of that institution. For example: abc@ugc.ac.in; abc@satyam.net etc.

13.4 TYPES OF EMAIL ACCOUNT

Based on access to email provided, there are two types of Email Account

13.4.1 POP/IMAP Account

A POP account is based on Post Office Protocol and supports "offline" Email management. In a POP account, when a user connects to a mail server through a mail client, the client retrieves all the messages from the server, stores them locally and marks them as new/unread messages. Subsequently, the downloaded messages get deleted from the server and the connection is closed. Most of the ISPs (Internet Service Providers) offer POP mail.

An IMAP account is based on Internet Message Access Protocol. It allows their users to work with their messages in both online and offline modes. In this, the email client retrieves the message headers from the server and can store local copies of the messages in a local (temporary) cache. All the messages are left on the server until the user deletes them. This mechanism allows multiple email clients to access a single mailbox and is often used for corporate / business emails (e.g. sales@company-domain.com).

13.4.2 Email Forwarder

This type of email accounts will forward any incoming mail to another email address. Normally, all the service providers give email forward service.

13.4.3 Mailing List

A Mailing list consists of its subscribers/members' email addresses. Any email sent to a mailing list account will be distributed to all the subscribers of the mailing list. For example, lis-forum@ncsi.iisc.ernet.in

13.4.4 Auto Responder

Within an email account, an autoresponder can be set with a readymade reply to any incoming email such as *message for successful receipt of mail*; *vacation email*, etc.

13.4.5 Email Bouncer

An email bouncer enables to send a 'fake' bounce message to the sender. Thus, the spammer is led to believe that the email account is inactive or unrecognised by the server, and in many cases will remove the email address from its mailing list.

13.4.6 Email Blackhole

To avoid spam mails from certain addresses, a blackhole for those addresses is created so as to avoid/discard any messages coming from those addresses.

Self Check Exercise

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

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

1) Discuss needs of email.

2) What is email forward?

.....

.....

.....

.....

13.5 STRUCTURE AND FEATURES OF EMAIL

Internet email messages consist of two major sections: Header and Body

13.5.1 Header

The header of an email is structured into various fields such as summary, sender, receiver, and other information about the email. The header of an email can be easily distinguished from the body of the email.

Various fields included within header are:

1) **From**: Contains the email address, and optionally name, of the sender of the message.

2) **To:** The email address(es), and optionally name(s), of the receiver(s) of the message.

3) **Subject:** A brief summary of the contents of the message.

4) **Date:** The local time and date when the message was originally sent.

5) **CC:** Stands for Carbon Copy. It contains email address(es) of those who will receive a copy of the message in addition to receiver(s) mentioned in To field.

6) **Message-ID:** It shows the number assigned to the message by the mail program at the host machine.

7) A series of **Received:** These are the lines, showing details of the systems through which the email has passed (useful for troubleshooting if mail bounces back).

8) A **Reply-to:** gives the preferred address for replies (usually but not always the same as the sender's address).

Note: Blank spaces are not allowed in an email address. Also, an email address is not case-sensitive.

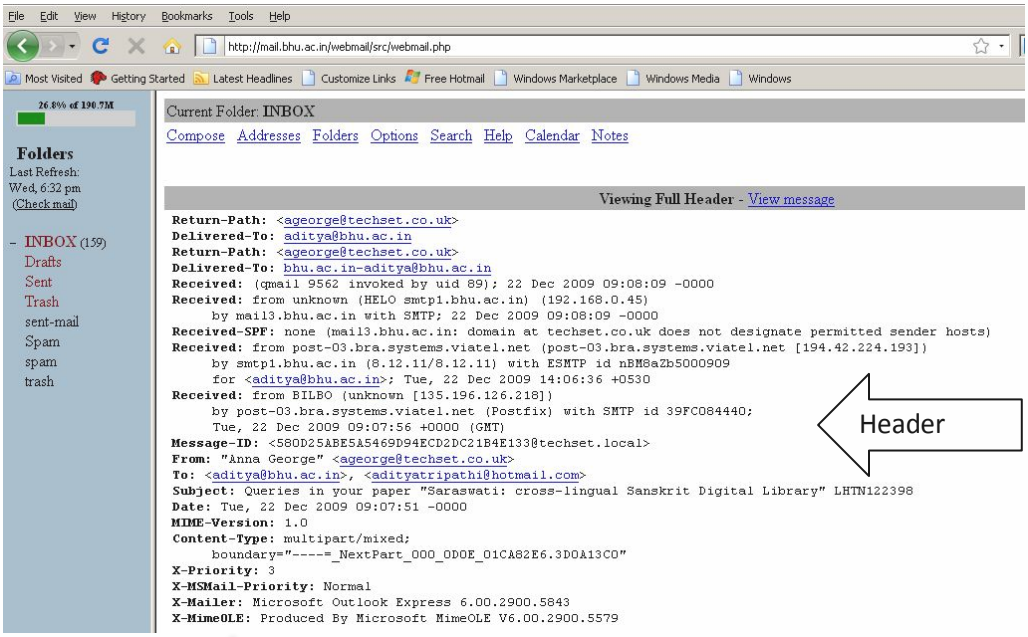

Fig 13.1: Header Section of an email

13.5.2 Body

The Body of an email contains the message itself mostly in text form. An account may be configured to automatically assign a signature (of the user) at the end. Signature is the text appearing at the end of the body by default in each message. Normally, it is the name of the sender and other contact details.

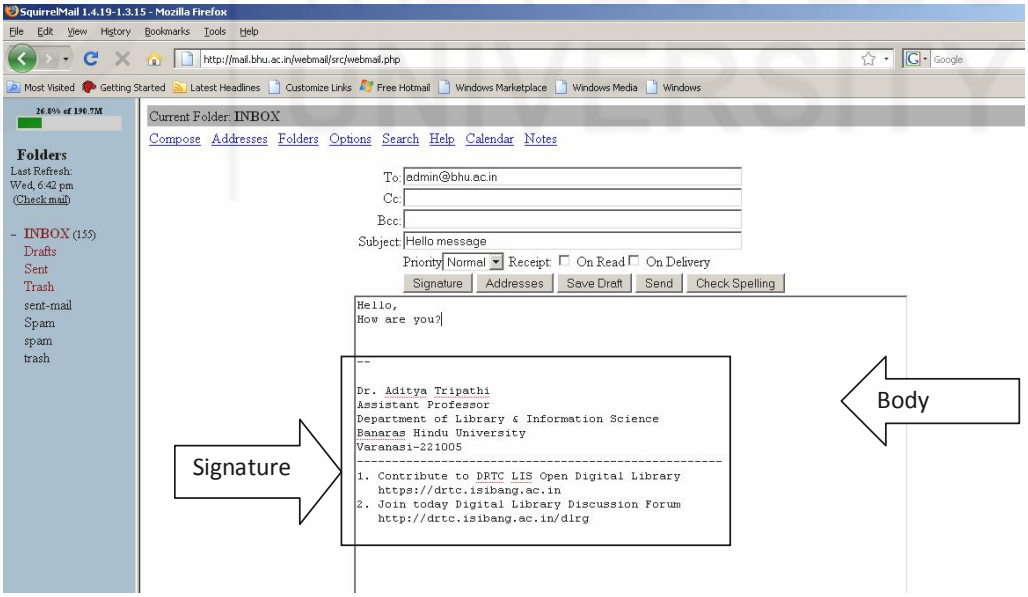

Fig13.2: Body Section of an email

13.5.3 Features

1) Email is based on push technology, i.e. Email is delivered to the recipient so they don't have to work to get it — they just need to open their Inbox to access the email.

2) Most of the email clients do offer to create MIME (Multipurpose Internet Mail Extensions) and HTML emails with colorful fonts, graphics and links.

3) The email account can be set to remind an upcoming event.

4) The accounts do offer facility for spell check while writing and email.

5) Any kind of document including multimedia objects can be sent through an email as attachments. However, the maximum size of the object, which can be attached or received, is fixed and set by the service providers. For example, MSN Hotmail can handle an object with 10 MB as attachment.

6) An email collates threads of communication on a single subject. Therefore, it sometimes acts as time stamped proof of communication. Gmail of Google offers this service.

7) A message can be saved/ printed along with all communication details.

8) A list of contacts (along with other details such as phone number, fax, etc.) can be created within an account with an ease to recall nickname associated with each email address. Hence, the user only has to enter one word for an email, instead of the full address.

9) Web-based email programs usually have a virus scan function that scans attachments before they are sent along with the main email.

10) One common feature of all email programs is the use of folders. These folders include an inbox, drafts folder, sent items folder and deleted messages folder. The users can also create other folders to better sort their email. Folders help categorise email according to its subject or importance. Filters are also included in email programs. These allow a user to define certain words or phrases that the program will look for in a message. The programs will then delete the message, forward it to a specified address, or put it in a particular folder.

Self-Check Exercise

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

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

3) What are the parts of an email?

.....

.....

.....

.....

13.6 FUNCTIONING OF EMAIL SYSTEMS

In this section we will know how an email system works.

13.6.1 Protocols

Protocol can be defined as a set of rules to perform a specific task. The mail server and its client exchange information with each other using a variety of protocols. A protocol is a standardised mechanism used at each end of a communication channel, to achieve

proper transmission of information. The most common of these email protocols is listed below:

IMAP Protocol

IMAP (Internet Message Access Protocol) is used by the client, which is used to read the email like Firebird, Outlook express, Apple Mail etc. (also called as Access Client). In this, email is received and held on Internet server for email account holder is read through a client. IMAP4 is the latest version in use. Only specific email messages requested by the user get downloaded from the server. It works well even for slower connection to Internet as it requires only a small data transfer. Also, one can create and manipulate folders or mailboxes on the server, make a keyword search through email, delete messages etc.

POP3 Protocol

The **POP (Post Office Protocol 3)** protocol provides a simple, standardised way for users to access their mailboxes and download messages on their computers using email clients.

This protocol permits to download all email messages from the mail server to a local computer. One can also choose to leave copies of their emails on the server as well. The advantage is that once the messages are downloaded mail can be read without an internet connection. Most current email client applications support POP3.

SMTP Protocol

The **SMTP (Simple Mail Transfer Protocol)** protocol is used by the Mail Transfer Agent (MTA) to deliver an email to the recipient's mail server. The SMTP protocol can only be used to send emails, not to receive them. The vast majority of mail servers use SMTP.

HTTP Protocol

The **HTTP protocol** is not a protocol dedicated for email communications, but it can be used for accessing mailbox. This protocol can be used to compose or retrieve emails from an email account. Hotmail, Yahoo are few examples which are using HTTP as an email protocol.

13.6.2 Delivery Agent

Mail Transfer Agent

Mail Transfer Agent is a piece of software which transfers messages or mails from one host or machine to other. It is often referred as mail server. MTA is the program which forwards message from one machine to other and finally delivers to the destination. An MTA receives a mail and puts a Received stamp in the header of the mail and checks for the recipient on the host. If the recipient is not present on the host MTA transfers the message to next host and thus passing through various hosts message is delivered to the recipient. The message is forwarded using SMTP. Example of MTA is qmail, sendmail, postfix etc.

Mail Delivery Agent

A mail delivery agent or Message Delivery Agent (MDA) is computer software used by the Mail Transfer Agent to deliver email to a particular user's mailbox. When the mail server receives a mail or in other words MTA (Mail Transfer Agent) using SMTP

than the mail is placed in the targeted user's mailbox. This work is done by MDA locally. Some of the mail delivery agents are: binmail, delivermail; maildrop, postfix-maildrop etc.

13.6.3 Access Client

The term *email access client* refers to an agent acting as a client towards an email server to access an email account. It is kind of application software. Some of the examples of Access Client are Outlook Express, Outlook or Thunderbird.

Note: Thunderbird is a free open source, cross-platform email client that has quickly gained huge popularity, with a reason. It is extremely easy to use, fairly powerful and robust.

13.6.4 Setting up Account

Several websites provide free email services. Following is an example for creating an email account on Yahoo! Mail.

1) Go to the website **https://mail.yahoo.com/** (Fig. 13.3). This provides a login page for the users who already have a mail account with Yahoo. For new users of Yahoo, they need to utilise **Sign Up**.

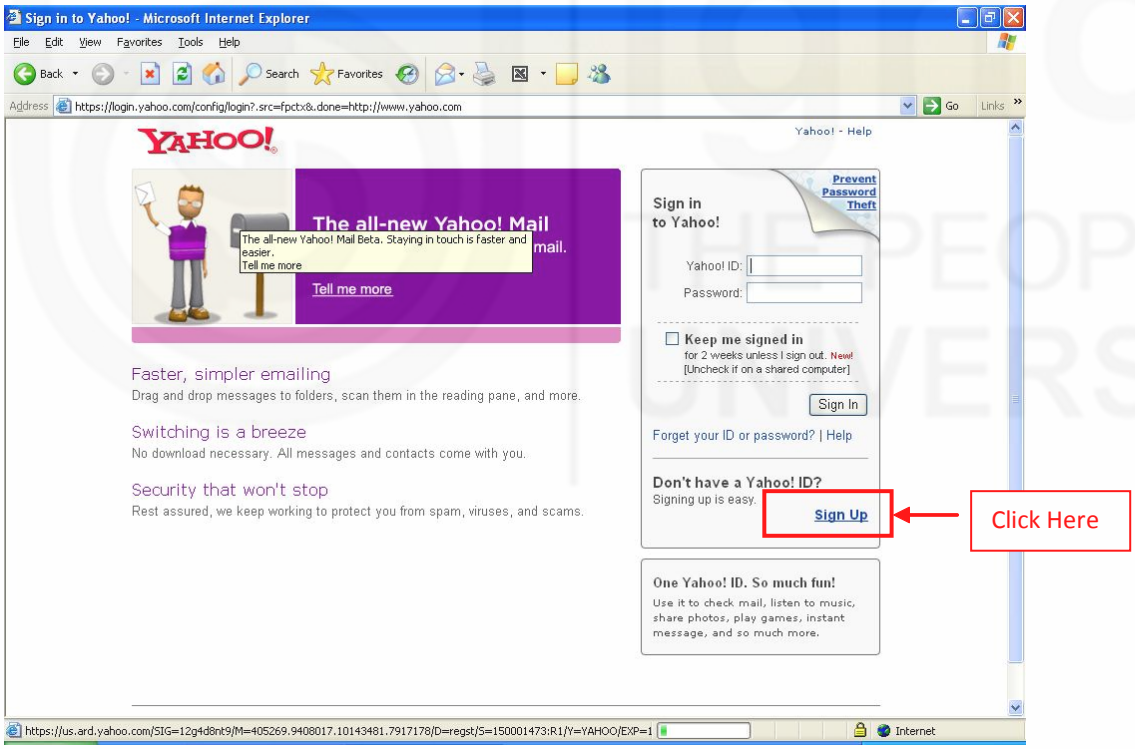

Fig 13.3: Yahoo Mail Homepage

2) Click on **Sign Up.** This will open a Registration Form (Fig. 13.4).

3) Enter the registration details and choose an ID and password for your Yahoo account. Availability of a Yahoo ID can be checked through the button Check . If an ID is unavailable, this will enlist other options for Yahoo ID.

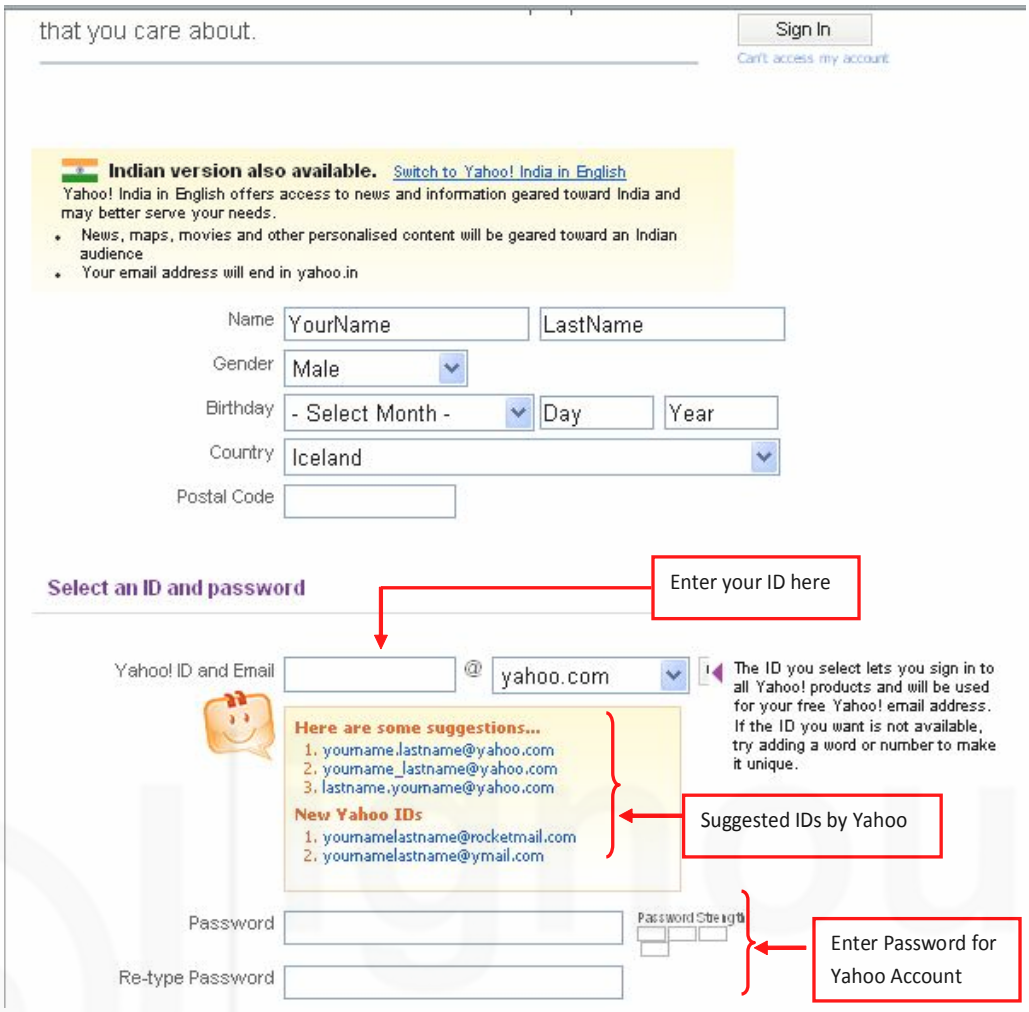

Fig. 13.4: Yahoo Mail registration Form (I)

4) Enter the verification code and click the button "Create My Account" (Figure 13.5).

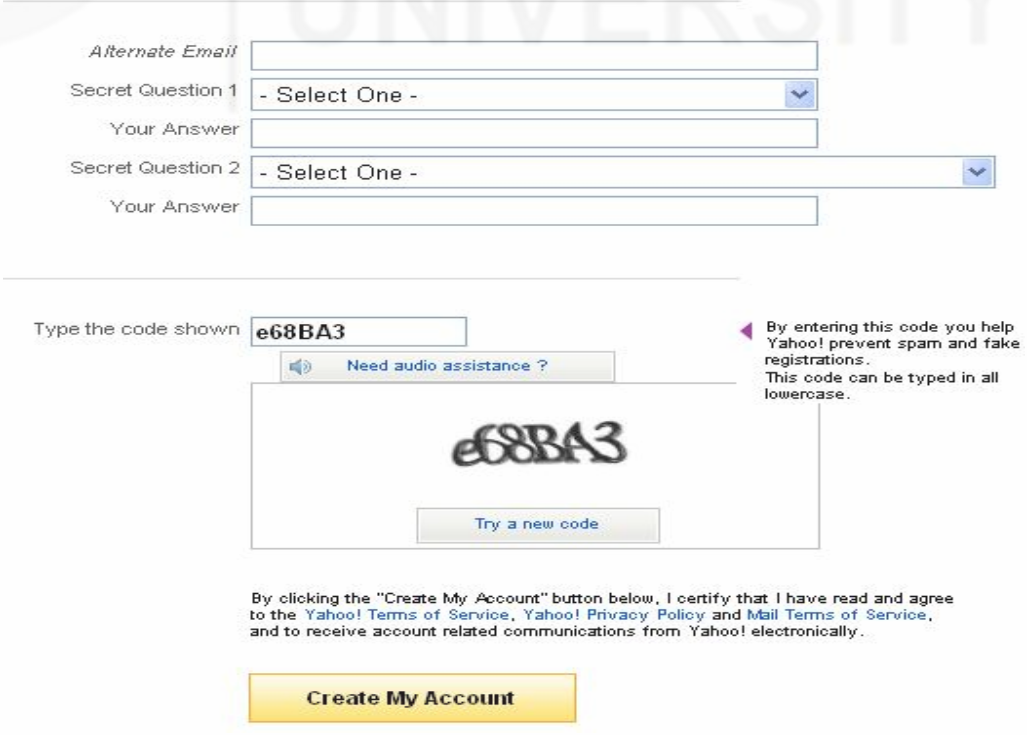

Fig. 13.5: Yahoo Mail registration Form (II)

5) This will open a new page confirming the creation of new yahoo account along with registration details Fig. 13.6. Click **Continue** to access the newly created account.

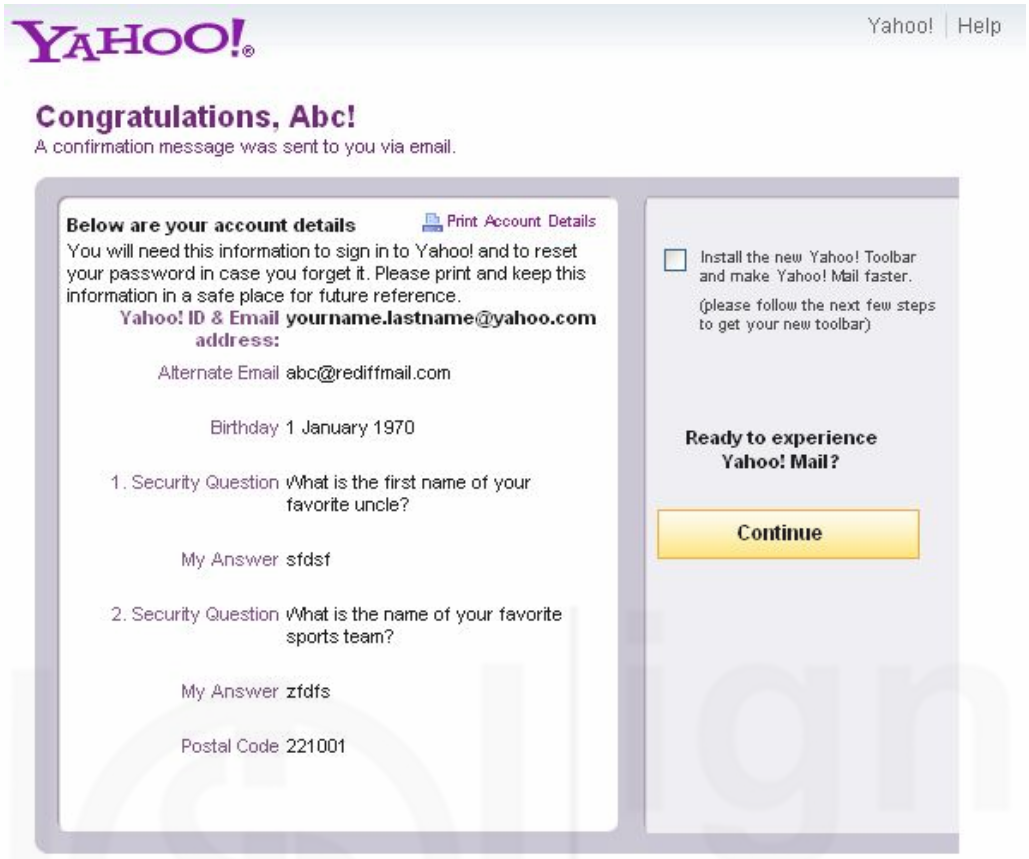

Fig. 13.6: Confirmation Message for new account

NOTE: To access the account for the next time, the Yahoo ID and password can be used on Sign-in page to access the mail account.

Most of the Web based Mail services offer four standard email folders: **Inbox**, **Sent**, **Drafts** and **Trash**, each of them are described below:

- **Inbox:**Lists all emails received from other email accounts. Highlights newly received/unread mails.
- **Sent:** A copies of messages sent are put into the **Sent** folder provided that the mail account is set to save all the sent messages.
- **Drafts:** A place for storing unfinished messages. If writing of a message is not yet finished and needs to be stopped in between, clicking the **Save** button puts the message into the **Drafts** folder, which can be accessed later on.
- **Trash:** stores email that is deleted from other folders. The messages are not truly deleted until they are deleted from this folder.

A message can be opened by clicking its **Subject**. Once the mail gets opened, at the top, options are available to delete, print or forward a message.

Note: Messages in Trash folder that are more than few days old will automatically be deleted.

13.6.5 Folder Management

Once a Yahoo account is established, one can create folders other than the already existing permanent folders (such as Inbox, Sent, Drafts, Trash and Spam). There is panel at the extreme right side of the panel called Folders (Figure 13.7).

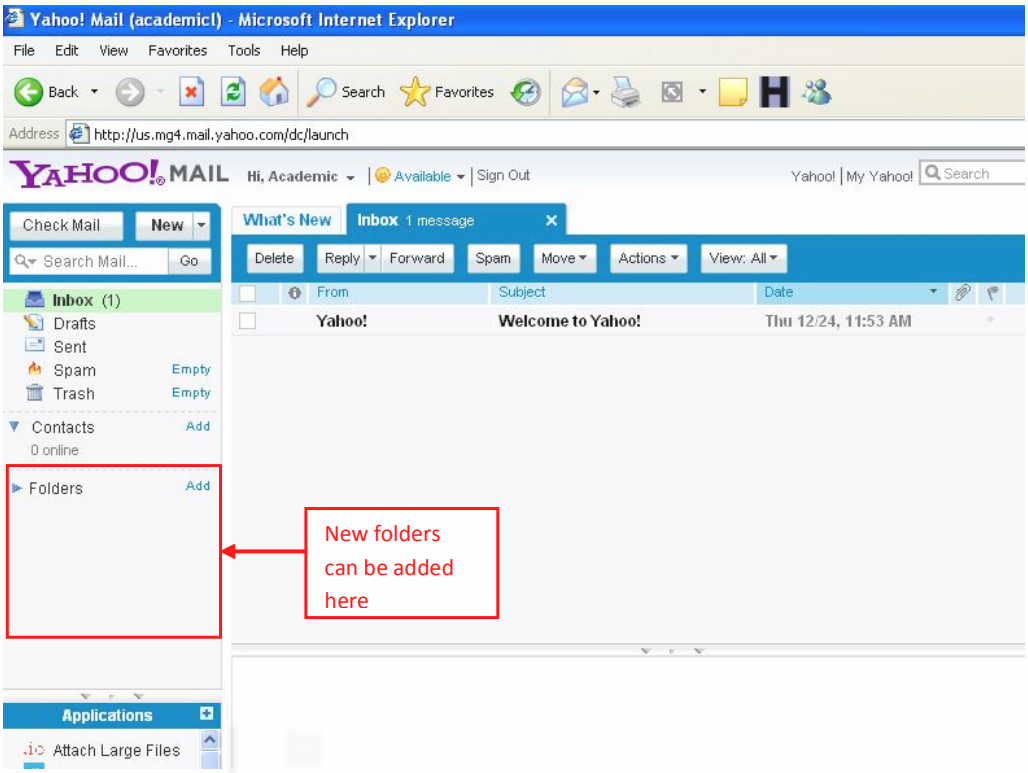

Fig. 13.7: The Folders Panel

To Create a New Folder

1) Click on **Add** (listed in front of the option **Folders**). This will create a new untitled folder below the panel Folders.

2) Type the name for the new folder e.g. FORUM (Fig. 13.8)

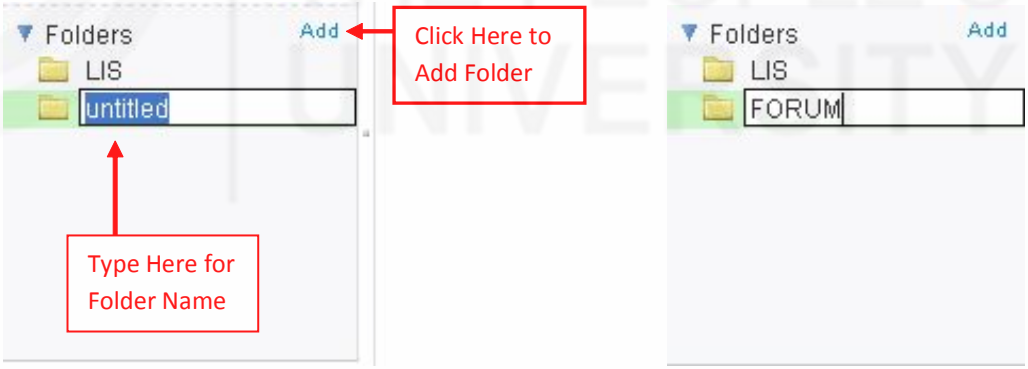

Fig. 13.8: To Create New Folder

To Rename a Folder

The steps are as follows (Fig. 13.9):

1) Select the folder to be renamed by clicking on it. Now, either press **F2** on keyboard or right click on mouse and select **Rename.**

2) Now, type the new name for the folder and press Enter from keyboard.

3) The folder will get assigned the new name.

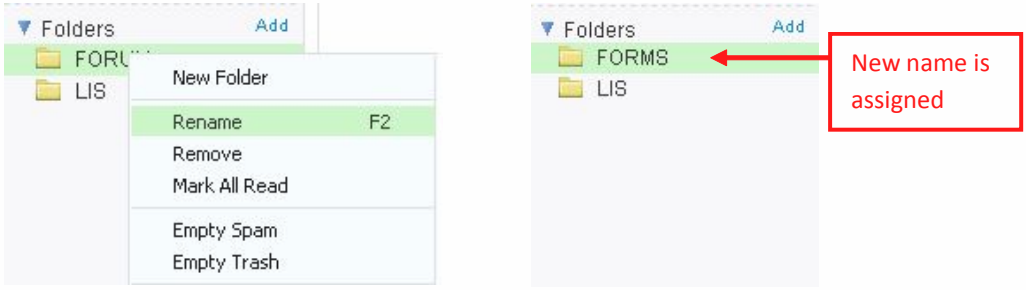

Fig. 13.9: To Rename a Folder

To Delete/Remove a Folder:

1) Select the target folder (e.g. LIS)

2) Delete all mails from the target folder.

NOTE: A folder can be deleted only if it is empty i.e., no mails are present. Hence, before deleting a folder each mail should be deleted from the folder.

3) Now, Press **F2** on keyboard or right click on mouse and select **Remove**. This will prompt a message to confirm the action. Click OK. (Fig. 13.10)

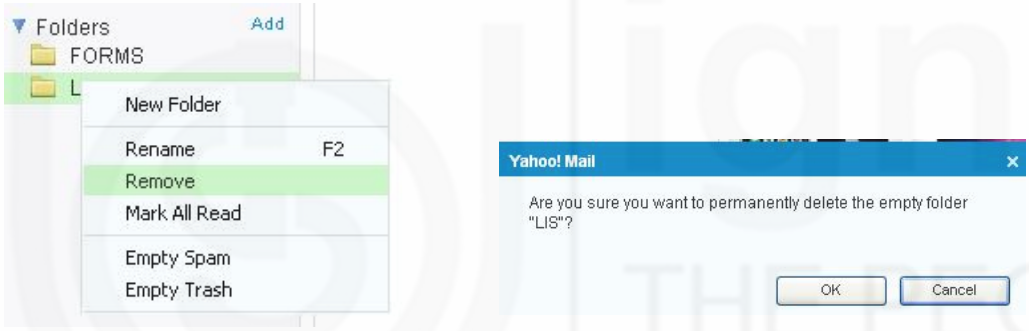

Fig. 13.10: Removing the LIS Folder

4) The folder is deleted successfully (Fig. 13.11).

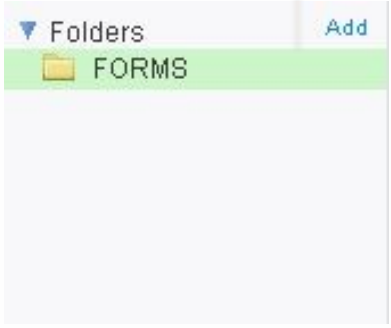

Fig. 13.11: LIS folder is deleted from the list

To Move an Email to a Folder

1) Go to the Inbox or Sent Folder.

2) Select the target email.

3) Click **Move** button of the mail task panel (Fig. 13.12). All the created folders (in this case FORMS and LIS) as well as Spam and Trash folders will be listed.

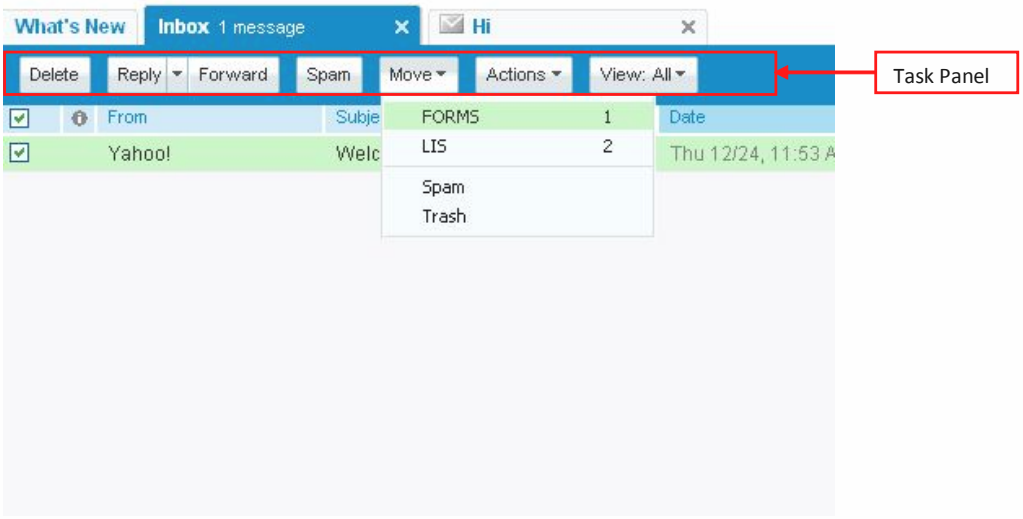

Fig. 13.12: The Task Panel of Inbox

4) Select the folder where the mail has to be moved (e.g. FORMS folder).

5) The mail is successfully moved to the target folder (Fig. 13.13)

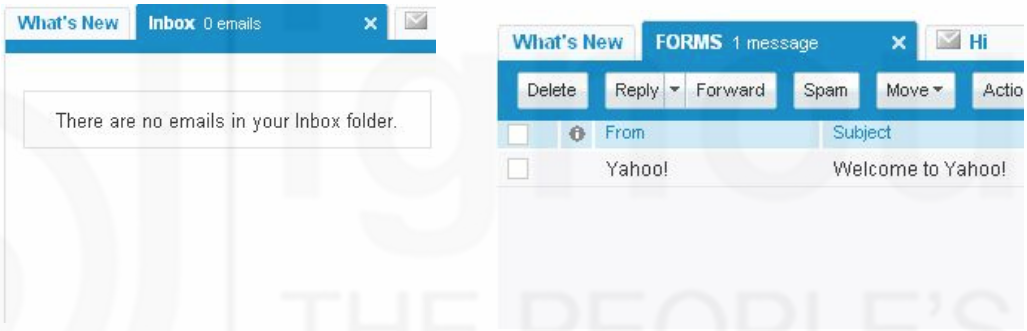

Fig. 13.13: The Empty Inbox and FORMS Folder with Message

Self-Check Exercise

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

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

4) What is Mail Transfer Agent?

.....

.....

.....

.....

13.7 MESSAGING

Messaging is a method of communication between two people or organisation. Messaging could be done using the power of Internet or through cell phones. Messaging software connects two people and facilitates them to communicate through text or voice or both of them.

Two types of Messaging

a) **Asynchronous Messaging**: The term "asynchronous messaging" means a method of communication between programs in which a program places a message on a message queue and leaves. It really does not bother how the message will be delivered. It is the delivery agent or the kind of infrastructure ensures delivery of message, even if the recipient is offline. For example, delivery of emails.

b) **Synchronous messaging**: In this kind of communication sender and receiver both have to be in connectivity while transferring the message. For example, telephonic conversation. Hence, a program places a message in a message queue and then waits for a reply to its message before resuming further.

13.7.1 Instant Messaging

Instant messaging (IM) is a form of real-time communication between two or more people based on typed text or using audio or video. The message is conveyed via devices connected over a network. Most IM programs provide these features:

- **Instant messages/Chat:** sending and receiving text/notes with an online friend.
- **Chat Rooms:** a common platform where two or more than two people can communicate.
- **Files/Web links/Videos/Images:** can be shared over network.
- **Talk:** Instead of a phone, Internet can be used to actually talk with friends. e.g Google+ Hangouts
- **Mobile capabilities:** Instant messages can be sent to mobile/cell phones.

13.7.2 Unified Messaging

Unified Messaging (or UM) is combination of different media into one channel. A user can access information into different media using a single device. Normally, Unified Messaging is common in mobile communication where, voice, text and fax can be accessed using one mailbox. It provides power to reach people almost anywhere, at any time and the flexibility to allow people to control when they can be reached.

13.8 ISSUES WITH MESSAGING

IM is increasing in popularity in both professional and personal applications. However, as with most of the things are Internet based, the increasing use of instant messaging has led to an associated increase in the number of security risks.

13.8.1 Spamming

Often in the email accounts one receives lot of unwanted mails which includes threats, promotional mails and so on. These messages are called Spam. There are two kinds of spams, mail lists and individual mails. Often users subscribe mail lists where lots of promotional materials are posted in bulk. These materials are of generic nature. There is another type of spamming where individual users are targeted for example, individual messages from unknown for illegal business or receipt of some unknown prize. Almost all the service providers have mechanism for filtering such messages based on the content or the subject line of the messages.

13.8.2 Privacy

Many Email service providers have their own privacy policies. Most of the free web based email services have financial dependency on advertisers and in return they share user information such as their email address information with advertisers. Also, details of the communication can be shared with government agencies on demand. Thus, one must be aware of the privacy policies of an email service provider before going for their services.

13.8.3 Security

Email /Instant messaging account are vulnerable to hijacking or spoofing. Spoofing is a phenomenon where an attacker hijacks another user's email/IM account and impersonates as user with others. Things apart the information exchanged over network can also be read over network by a third party. Hence, it is important to ensure information exchange.

There are several ways which have been devised to overcome all these problems which are enlisted below:

1) One of the best ways to secure the information being transmitted along an IM network is to encrypt it.

2) Keeping message logs for tracing any kind of mischief.

3) If file transfer via the instant messaging network is not required, then an instant messaging system that does not allow for files to be transferred should be utilised.

Self-Check Exercise

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

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

5) What is a Spam?

.....

.....

.....

.....

13.9 WIDGETS AND UTILITIES

Widgets are small programming code that users can add to their webpage, personalised homepage, web browser, desktop, blog or social network. Mostly, these codes are embedded within an image file, hence, can be evoked through a mouse click or through a keyboard command by a computer or Internet user. A widget is used to enhance the look and feel of a website i.e., to make it look more fanciful. Widgets are considered as an offering by Web 2.0 to internet users.

There are two kinds of widgets

a) **Web Widget**

They are web-based, which means they can only be part of a website. It cannot be put on desktop unless customised to be used as desktop widget. The best example of a

web widget is a YouTube video being displayed on a website other than YouTube. Similarly, *AddThis* (http://www.addthis.com) is a simple free web widget which can be used to bookmark a particular piece of information through various social networking websites such as Delicious (http://delicious.com/), Stumbleupon (http://www.stumbleupon.com), Google Bookmarks.

b) **Desktop Widget**

These widgets need not be part of a website and could be integrated on the desktop of a computer. These application accesses the Internet to provide the latest information. For example, a widget "Desktop Weather" displays the temperature and forecast for a given (configured) geographical area through the website **Weather.com.**

Fig. 13.14 : Web Widget

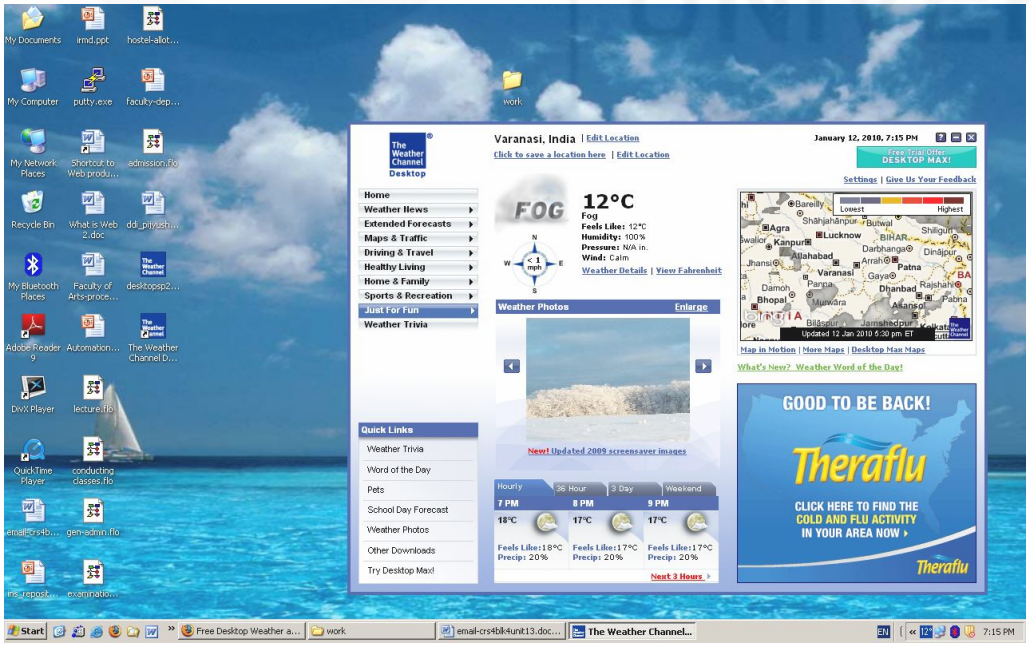

Fig. 13.15: Desktop Widget

Widgets can be delivered as:

- **Graphics:** Used to create banners, unusual fonts and animated images or text.
- **Gadgets:** hit counters, weather forecasts, calendars and maps
- **Entertainment:** Streaming of TV listings, daily quotes, interesting facts or games
- **Social:** provides link to social bookmarking or networking pages such as Facebook or Twitter e.g. Addthis.com
- **Audio Visual:** e.g. Streaming of YouTube videos directly to a web page

Utilities:

Utilities are useful for fixing minor problems or mis-configurations and handle day-to-day chores associated with computers. They are designed to help manage and tune the computer hardware, operating system or application software by performing a single task or a small range of tasks. Hence, the objective is to keep a system running at peak performance. It is also known as service *program*, service routine, tool, or *utility* routine.

Some of the examples of mail utilities are:

- **Cryptographic** utilities encrypt streams and files;
- **Mail Send Utility** to send multiple mails to a desired mail list from command line;
- **Email Finder Pro** n/a Fast and simple email address extraction utility;
- **JPEE Email Utility Lite 5.3.2** is known as a tool to merge and extract email, parsing data. JPEE is available as a free, easy to use, highly configurable, email communications software implementation;
- **Google Email Uploader** is a useful and smart desktop utility for Windows with the ability to upload email and contacts from other desktop email programs (like Microsoft Outlook) into your Google Apps mailbox. The Email Uploader preserves information such as sent dates and sender/recipient data, as well as the folder structure used by the other email program.

13.10 SUMMARY

Email is the most used service on Internet. It has become the fastest medium for transferring the message as well as data. The importance and authenticity of email is now being recognised officially and many organisation use email for day-to-day communication and do broadcast there circulars and orders by email. The service is offered by several service providers free of cost however, one can set his own mail server for communication. In this chapter we have learnt about the email systems and different types of email addresses. In due course we have also learnt about Instant Messaging services. There are some issues like individual privacy and security which has been discussed in detail in order to give insight of the problems.

13.11 ANSWERS TO SELF CHECK EXERCISES

1) Email is the fastest method of communication over Internet. Email stands for electronic mail. An email is delivered within no time irrespective to the distance. It carries text, audio, video and graphics as an attachment. Using email one can

transfer any piece of information or data to a number of people in a single stroke. It is one of the most secure medium for one to one as well as one to many communication.

2) Forward is a facility normally provided with all the email services. If one wants to change his email address to a new one and wants to receive mails from the previous accounts in the newer one he/she can put a forward on the previous accounts which will forward all emails received on the previous accounts to the new account.

3) An email has two parts:

   - Header
   - Body

   Header contains information regarding whom the mail is sent and email address of sender with brief subject about the content of message and the date mail is sent. The header also shows email addresses of all those who have received a copy of the mail.

   Body contains the message from the sender.

4) Mail Transfer Agent (MTA) is a piece of software, which transfers messages or mails from one host or machine to other. MTA forwards message through different machines to the destination. A mail could be transferred through several machines or MTAs and each machine stamps over email.

5) Spam is unwanted emails received in the inbox. Most of the such mails are propaganda materials. Apart from this these mails may contain harmful programs or virus.

13.12 KEYWORDS

Attachments : Files, which can be sent along with email.

Blacklist : This is a list of email addresses, which are blocked for sending mails. In other words, mail server refuses to accept mails from blacklisted email addresses.

Blind Carbon Copy (BCC) : Copy of an email message sent to recipients but his email address does not appear in the message.

Bounce : Returned messages that do not reach to destinations.

Carbon Copy (CC) : Carbon copy is a copy of an email sent to addresses other than the main recipient.

DNS : Stands for Domain Name Server, which translates domain names into IP addresses.

Download Message : Allows saving a copy of email message on local computer

Email harvesting : A program to scan Web pages for the purpose of collecting email addresses.

Email headers : Section of an email message contains the sender and recipient's email addresses and routing information.

Filters : Filters automatically filter incoming emails into designated folders. Filtering is done based on different fields of an email like, recipient address, subject, domain name and so on.

Forwards : Feature of an service which automatically forwards incoming emails to another email account.

MIME (Multi-Purpose Internet Mail Extensions) : An extension of the email standard that allows users to exchange multimedia files.

Plain text : Text in an email message that contains no formatting elements (i.e., color, bold, italics, etc.), pictures, or HTML.

Post Office Protocol (POP3) : A standard that enables emails to be retrieved from a remote mailbox. It allows to collect emails from an account to an email client program like, Outlook.

Proxy : A computer system or router that breaks the connection between sender and receiver, giving anonymity.

Reminders : Added feature of email systems that allows setting up reminders on a certain date.

Search Facility : A feature within the email system to search an email based on keywords/phrase as well as on certain criteria.

Signature : Section in the email message which goes by default with the sent email normally carrying name, address, website, address, phone number, etc. It appears at the end of the message.

SMTP : A protocol standard that enables emails to be transferred from sender machine to destination machine.

SMTP(Simple Mail Transfer Protocol) : Stands for Simple Mail Transfer Protocol, used to send email on the Internet.

Snail mail : Postal service delivers letters at home in paper format via postman.

Soft bounces : Bounce of a mail with a message due to temporary network congestion.

Spam : Unsolicited, unwelcome email sent in a large volumes.

Spell Check : Feature with email service to correct any spelling errors.

Spoofing : A method of spamming where the spammer uses false email address which does not exist.

Web mail : Web browser based email client like squirrel mail.

White list : A list authorised email addresses from which email messages can be delivered regardless of spam filters.

13.13 REFERENCES AND FURTHER READING

1. Email Protocols: IMAP, POP3, SMTP and HTTP. <http://www.emailaddressmanager.com/tips/protocol.html>
2. Using Folders. <http://www.earthlink.net/webmail/help/earthlink/en_US/using_folders.html>
3. What are utility programs. <http://depts.alverno.edu/cil/mod1/software/utility.html>
4. Widgets. <http://www.mediafuturist.com/files/widgets_trendmarker_feb08_forweb_200803251313542.pdf>
5. Electronic mail. http://tldp.org/LDP/nag/node186.html
6. Unified messaging. <http://www.iec.org/online/tutorials/unified_mess/topic02.html>
7. Email. (n.d.). Retrieved from Wikipedia. <http://en.wikipedia.org/wiki/Email>
8. Multipurpose Internet Mail Extensions. (n.d.). Retrieved from Wikipedia. <http://en.wikipedia.org/wiki/Multipurpose_Internet_Mail_Extensions>
9. Simple Mail Transfer Protocol. (n.d.). Retrieved from Wikipedia: <http://en.wikipedia.org/wiki/Simple_Mail_Transfer_Protocol>
10. Learn The Net: Email. (n.d.). Retrieved from Learn The Net: <http://www.learnthenet.com/english/section/email.html>
11. The history of email. (n.d.). Retrieved from Net History: <http://www.nethistory.info/History%20of%20the%20Internet/email.html>
12. Levine, J. R., Young, M. L. & Baroudi, C. *The Internet for Dummies*. 11th ed. Hoboken: Wiley Publishing, 2007. Print
13. Gralla, P. *How the Internet works*. 4th ed. Indianapolis: Que Publishing,1998. Print
14. Thomas, Brian J. *The Internet for scientists and engineers: online tools and resources*. 3rd ed. Oxford: Oxford University Press, 1997. Print
15. Lindsell-Roberts, S. *135 tips on email and instant messages: plus blogs, chatrooms, and texting.* Boston: Houghton Mifflin Company, 2008. Print
16. Wood, D. *Programming Internet Email*. Sebastopol: O'Reilly & Associates, 1999. Print

UNIT 14 WORLD WIDE WEB

Structure

14.0 OBJECTIVES

After reading this Unit, you will be able to:

- discuss the growth and development of WWW;
- explain the underlying technology behind WWW;
- describe Web 2.0 technology; and
- understand services of Web 2.0 technology and its impact.

14.1 INTRODUCTION

Internet has changed the life of people. While traveling in the train, airplane or any other mode of transportation one can keep oneself busy and connected to the world. It has influenced education, commerce, governance and entertainment, which have tremendous impact on day-to-day life of an individual. Internet was there even before twenty years but it is WWW (World Wide Web), which has brought radical change in the use of Internet. World Wide Web consists of interlinked hypertext documents. The transfer medium is known as Hypermedia, which carries the multimedia objects such as images, as well as audio and video files over network in addition to text. There is no doubt about the fact that the actual influence of Internet on society could be seen only after the introduction of WWW. It was a revolutionary break-through in the emerging technological environment.

14.2 WORLD WIDE WEB

The credit of developing World Wide Web (WWW) goes to Sir Tim Berners-Lee in 1989, and later on Robert Cailliau in 1990 at CERN laboratory in Switzerland where it was demonstrated over distributed hypermedia servers (ý1). The servers store hypertext documents, which can be accessed via a client (i.e., a web browser). The language used for creation of document is known as HyperText MarkUp Language (HTML). These web documents or hypertext documents are linked to each other using a specific pointer system known as Uniform Resource Locator (URL). These pointers are like *handles* and have the capability to fetch the document stored and scattered across various web servers. The servers where these hypertext objects are stored host the hypermedia documents, honouring the request to serve the documents to a client. Internet is collection of such servers also known as Web servers placed in different parts of world. Hence, WWW is a service over Internet. It is a collection of documents available over host servers worldwide.

A special kind of application software is used to access hypertext documents know as Web Browser. There are many web browsers available like, Mozilla Firefox, Internet Explorer, Opera, and so on. WWW supports hosting, dissemination and playing of multimedia documents that includes audio video.

In 1989, Sir Tim Berners Lee wrote a proposal "Information Management: A Proposal" for conceptualising WWW at CERN Lab. In 1991 the first web browser was released. The web browser used to work on hypertext with Graphical User Interface on platform called as NeXTStep. The web browser was a WYSIWYG (What you see is what you get) type with facility to support hyperlinks (Fig. 14.1).

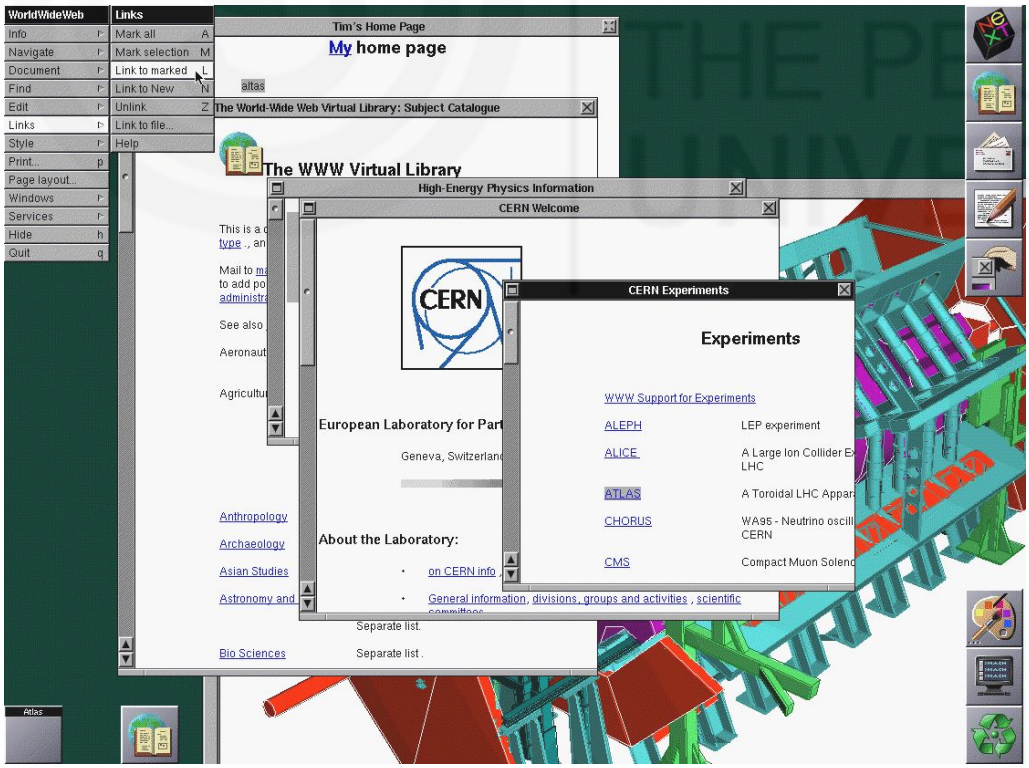

Fig. 14.1: Early Web Browser

14.3 CONCEPTUAL FRAMEWORK OF WWW

Web is one of the most widely used components of Internet. The Web allows access to information dispersed all across the world on different servers. The information is

available in diversified form such as text, graphics, animation, photos, audio and video. The Web physically consists of:

a) a personal computer or mobile device;

b) web browser software;

c) connection to an Internet service provider;

d) computers called servers to host digital data and routers and switches that direct the flow of information.

14.3.1 Communication Architecture

The basic web structure of WWW is two tiered popularly known as Client Server Model (Fig. 14.2). Those machines that provide services (like Web servers or FTP servers) to other machines are **servers**; and the machines utilising those services are known as **clients**. This kind of architecture depends upon the following:

a) **Standard Representation of Information on Web:** Markup Languages such as HTML, XML are some of the standards available for the content representation over the Web.

b) **Transfer Protocols:** These are the different protocols for transferring information between computers on the Internet. HyperText Transfer Protocol is the underlying protocol used by web. HTTP is a synchronous request-reply protocol that requires direct, online connections. HTTP defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands. For example, when you enter a URL in your browser, this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Web page. HTTP is called a *stateless* protocol because each command is executed independently, without any knowledge of the commands that came before it.

c) **Addressing Protocols:** Web utilises protocols to identify a web object based on names and addresses. These are known as Uniform Resource Identifiers (URIs). Uniform Resource Locator (URL) is an example of URI. It is the global address of web documents and objects. The first part of the address is called as protocol identifier and it indicates what protocol to use, the second part is called as resource name and it specifies the IP address or the domain name where the resource is located. The protocol identifier and the resource name are separated by a colon and two forward slashes. An example of addressing and addressing protocol is as follows:

http://www.bbc.co.uk/news/world-latin-america-11617094

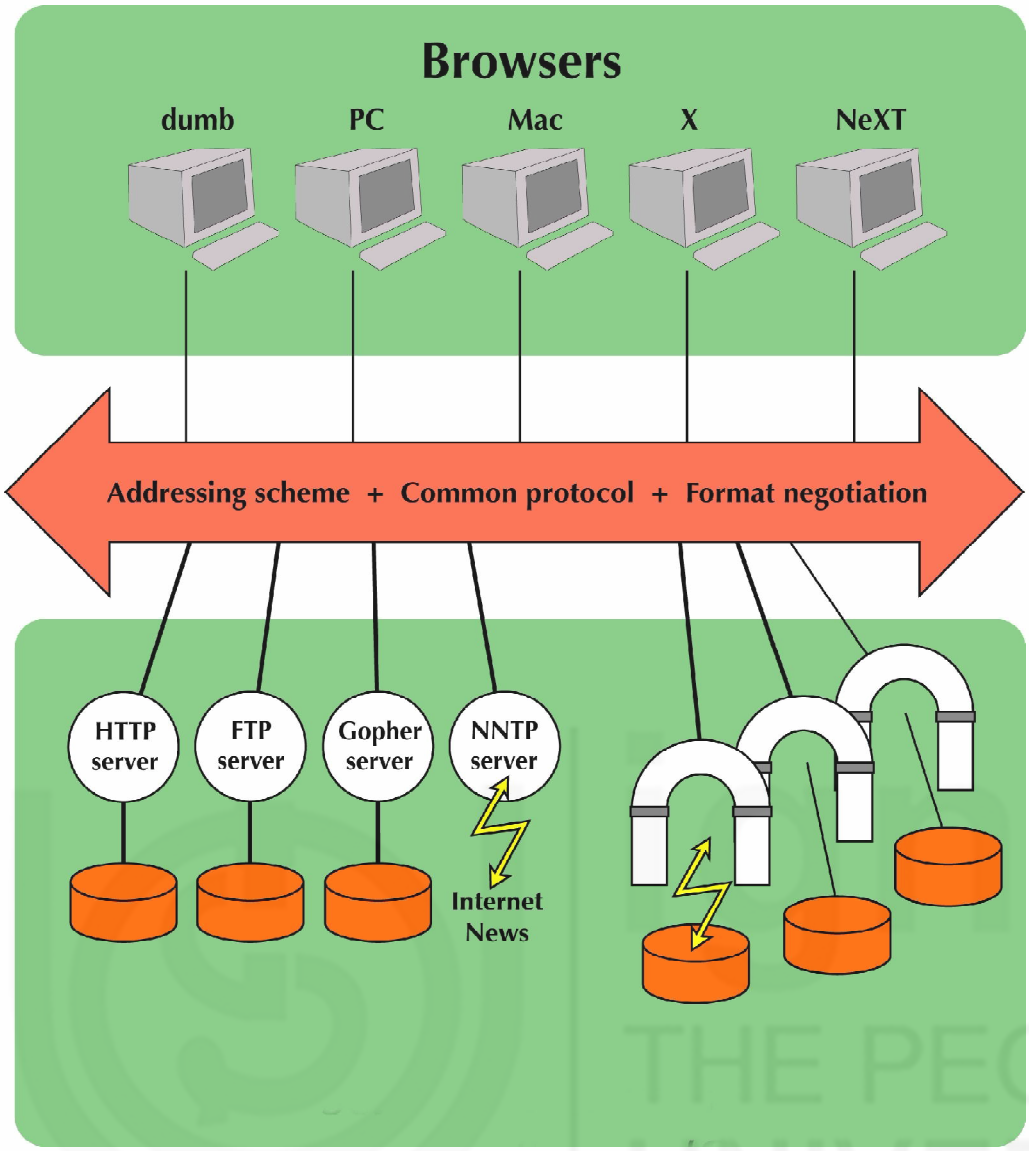

Fig. 14.2: Early Architecture of WWW given by Tim Berners Lee and Robert Cailliau

Self-Check Exercise

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

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

1) What is a Web Browser?

2) What do you mean by addressing protocol?

.....

.....

.....

.....

14.3.2 Protocols

Tim Berners-Lee implemented the HTTP protocol in 1991 at CERN, the European Center for High-Energy Physics in Geneva, Switzerland. HTTP stands at the very core of the World Wide Web. According to the HTTP 1.0 specification (26):

"The Hypertext Transfer Protocol (HTTP) is an application-level protocol with the lightness and speed necessary for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object-oriented protocol, which can be used for many tasks, such as name servers and distributed object management systems, through extension of its request methods (commands). A feature of HTTP is the typing and negotiation of data representation, allowing systems to be built independently of the data being transferred".

- HTTP provides a comprehensive addressing scheme with the concept of URI as a location URL. HTTP based hyperlink address is rendered in following format:

  http://host:port-number/path/file.html

  http://drtc.isibang.ac.in:80/

- HTTP protocol is based on a request from a client and response from a server. A server hosts web pages and on request, delivers these web pages to client. The default port from any HTTP request is 80, however one can change the port number to any other port number as desired. However, one port number can be used only to run one program. This kind of request/response format of communication is known as Client – Server Architecture. Every request is treated as a new request and there is no dedicated connection between client and server. This is known as stateless connection.

- HTTP supports multimedia objects also known as Internet Media Types (MIME format). The header of information informs the client what kind of data will follow.

14.4 MARKUP LANGUAGES

It is an encoding system to annotate the text of a web document. The system consists of notations known as tags. They control the structure, formatting as well as the relationship between different parts of a document. A marked-up document thus contains two types of text: text to be displayed and markup language on how to display it.

Example of markup language is HyperText Markup Language (HTML), one of the document formatting languages of the World Wide Web.

A marked-up document will typically look like this:

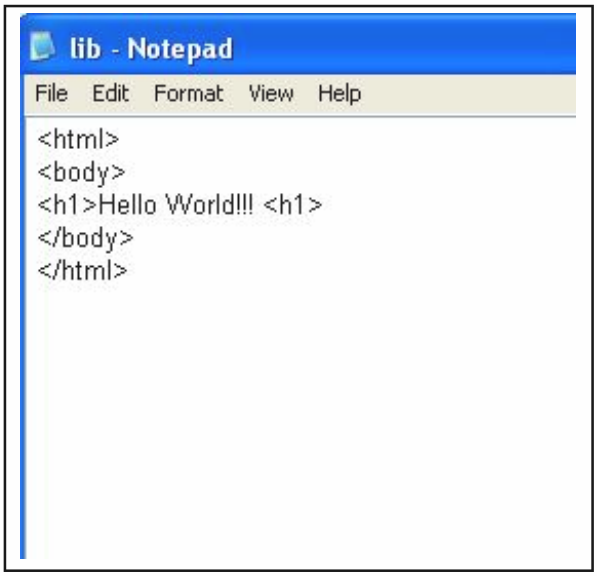

Fig. 14.3: lib.html (view through the notepad application)

The output of this document will be:

Fig.14. 4: lib.html (output view through Mozilla Firefox)

14.4.1 Definition and Need

A markup language is a methodology to annotate text for its representation and processing over World Wide Web. IBM researcher Charles Goldfarb is considered as "father" of markup languages. A markup language uses codes, which are also known as tags to describe the layout and formatting of the document. Also, it can also describe the types of information conveyed by a given text.

For example:

<html>

<Address> Flat No. 20, Rajajipuram, Lucknow </Address>

</html>

Thus, the piece of content included within <Address> tag is an address or a location. Over web, search engines processes these codes or tags and interpret the information accordingly. Thus, markup languages are important to define the meaning and context of the text on web. It also facilitates standard representation of text by different browsers.

14.4.2 Types of Markup Languages

Standardised Generalised Markup Language (SGML)

The markup languages that carry the instruction for text processing are known as 'Procedural Markup'. The idea of markup was to format a particular kind of document. But later on it was felt that markup languages could be used for system-to-system information interchange also. This was first realised by Charles Goldfarb, Ed Mosher and Ray Lorie when they were working with legal documents. They designed the first markup language known as GML (Generalised Markup Language) based on the following observation:

- The document processing programs needed to support a common document format.
- The common format needed to be specific to their domain - for example legal documents.

- To achieve a high degree of reliability, the document format would have to follow specific rules.

For example, take an example of a memorandum:

> To: Bishwanath Dutta
>
> CC: Bibhuti Bhushan Sahoo
>
> From: Aditya Tripathi
>
> Date: 27.01.2003
>
> Subject: Appointment order
>
> ---
>
> We are extremely happy to inform you that you are selected as the Coordinator of Knowledge Management Team.

There are six fields in this document:

- Who the document is intended for (the **To**: field).
- Who has been sent a copy of document (the **CC**: field).
- Who sent the document (the **From**: field).
- The date of document written (the **Date:** field).
- The subject of document (the **Subject:** field).
- The document body.

The structure of this document is fixed and one is bound to write it in the same structure. Hence, porting the information across systems will not be a problem as the structure of document is always same. The definition of the structure of document is known as 'DTD (Document Type Definition)'.

Once GML was designed, Goldfarb fine-tuned his work and proposed the SGML (Standardised Generalised Markup Language) which was further approved by ISO (International Organisation for Standardisation) in 1986. Hence, SGML is for defining the format in a text document. Readable by both humans and computer programs, SGML is usable in a wide range of applications such as print publishing, CD-ROMs, and database systems. SGML was not a language itself but it was a meta language to develop other markup languages. HTML (Hypertext Markup Language) is a derivative of SGML. HTML is more like a formatting language. Thus it is difficult to pull out what kind of data is stored inside a HTML document. Once this difficulty was understood, the need for domain specific tags was felt, for information interchange, which is not possible with HTML. Hence the XML was developed. It is always said that XML is more near to SGML when compared to HTML.

HyperText Markup Language (HTML)

HTML stands for HyperText Markup Language. It is a language, which is used to develop web pages. It is a collection of several tags to describe visuals of a webpage. The goal of HTML is to provide a display format to the given set of data so that it can be read on a web browser. HTML was originally designed by Sir Tim Berners-Lee in 1991 at CERN Lab. HTML is an offshoot of SGML.

HTML is the building block of a website. It allows multimedia objects to be embedded in the webpage including audio, video, text and graphics. The language consists of

tags. A tag is an element (known as HTML element), which has certain properties. These properties are applied on the data embedded in between tags. It is an individual component of an HTML document. Hence, HTML documents are collection of tags. These tags may simply contain data or can co-exist with other tags establishing parent-child relation. A tag has certain attributes, which are applied, on the contained data or on the child tags (or elements).

The number of tags used in HTML is fixed hence the language uses closed vocabulary. The structure of a web page is as follows:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
 "http://www.w3.org/TR/html4/strict.dtd">
<HTML>
 <HEAD>
  <TITLE>My first HTML document</TITLE>
 </HEAD>
 <BODY>
  <P>Hello world!
 </BODY>
</HTML>
```

Structure of HTML Document:

An HTML document has two parts,

- Head, and
- Body

'HEAD', contains elements (tags) for TITLE of the document. The 'TITLE' element stores information about the title of the document.

```
<TITLE>Website of Indira Gandhi National Open University</TITLE>
```

There is another element used in 'HEAD' section i.e., 'META' element. 'META' element stores information about the document such as author, copyright, location, relation, keywords and so on.

```
<META name="Author" content="Sneha Tripathi">
```

In the META tag first attribute or the property is defined under the NAME attribute and its value is given under CONTENT. In the above example Author is an attribute and value of author is 'Sneha Tripathi'.

These attributes are the attributes of the document, which is being described. Sometimes an attribute may use closed vocabulary or a scheme. In such cases META element also specifies the SCHEME used.

```
<META scheme="ISBN" name="identifier" content="0-8230-2355-9">
```

The second part of an HTML document is BODY element. The body of a document contains the document's content. The BODY element contains all the tags or elements, which are used to display the data over a web browser. It includes a variety of tags such as

```
<H1>…</H1>
```

<H2>…</H2>
<H3>…</H3>
<H4>…</H5>
<TABLE>…</TABLE>
<P></P>
<B></B>
<I></I>

Each element inside BODY tag can have various attributes, which is defined in HTML Standard Specification. The current version of HTML specification is 4.01. The next version of HTML, which is due, is HTML 5.0.

Extensible Markup Language (XML)

According to the abstract from the XML Specification version 1:

"The extensible Markup Language (XML) is a subset of SGML that is completely described in this document. Its 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."

- XML stands for eXtensible Markup Language.
- XML is a markup language much like HTML.
- XML was designed to describe data.
- XML tags are not predefined in XML unlike HTML where the tags are pre-defined.
- XML uses a DTD (Document Type Definition) to describe the data.
- XML with a DTD is designed to be self-descriptive.

Following are the goals kept in mind while developing the specification for XML:

i) XML shall be straightforwardly usable over the Internet.

ii) XML shall be compatible with SGML.

iii) It shall be easy to write programs to process XML files.

iv) The processors could read the XML document easily.

v) XML document should be human-legible and reasonably clear.

vi) The XML design should be prepared quickly.

vii) The design of XML should be formal and concise.

viii) XML document shall be easy to create.

ix) Terseness in XML is of minimum importance.

XML is different from HTML in the following ways:

i) XML was designed to carry data.

ii) XML is not a replacement for HTML.

iii) XML was designed to describe data and to focus on what data is.

iv) HTML was designed to display data and to focus on how data looks.

v) HTML is about displaying information. XML is about describing information.

What can be done with XML?

i) XML does not DO Anything

XML was not designed to DO anything. Maybe it is a little hard to understand, XML was not developed to DO anything. XML is created as a way to structure, store and send information.

```
<?xml version="1.0" encoding="UTF-8" ?>
<book>
    <title>Application of expert systems in libraries and information centres</
title>
     <author>
      <f_name>Anne</f_name>
      <l_name>Morris</l_name>
     </author>
    <edition>1st Edition</edition>
     <place>London</place>
     <publisher>Bowker-Saur</publisher>
     <physical_desc>241 p.</physical_desc>
</book>
```

Fig. 14.5: An XML Document in a Web Browser

The example shows the structure of a document, which describes a book, titled 'Application of expert systems in libraries and information centres'. The book has a title, author, edition, publication related information, etc. Author Name is further divided into first name (f_name) and last name (l_name). Inside these tags the actual data is stored. If one browses the document in the browser, data will appear embedded in the tag without having any kind of formatting.

ii) Define Your Own Tags

XML provides the facility to create domain specific tag set which facilitates the information interchange within a specific domain. For example, NewsML is developed for information interchange among the news agencies like Reuter and others.

iii) XML is Not a Replacement for HTML

It is important to understand that XML is not a replacement for HTML. It is useful for describing the data.

iv) XML can be used to Exchange Data

With XML, data can be exchanged between incompatible systems. In the real world, computer systems and databases contain data in incompatible formats. One of the most time –consuming challenges for developers has been to exchange data between

such systems over the Internet. Converting the data to XML can greatly reduce this complexity and create data that can be read by many different types of applications.

v) XML can be Used to Share Data

With XML, plain text files can be used to share data. Since XML data is stored in plain text format, hence it provides a software as well as hardware-independent way of sharing data.

This makes it much easier to create data that different applications can work with. It also makes it easier to expand or upgrade a system to new operating systems, servers, applications, and new browsers.

vi) XML can Make Data More Useful

With XML, data is available to more users. Since XML is independent of hardware, software and application, one can make their data available to more than only standard HTML browsers.

Other clients and applications can access XML files as data sources, like they are accessing databases. The data can be made available to all kinds of "reading machines" (agents). For example, a data set can be used to see a webpage in a web browser of a computer or it can be used see the display in mobile phone.

vii) XML is used to Create New Languages

XML is the mother of WAP (Wireless Application Protocol) and WML (Wireless Markup Language). The Wireless Markup Language (WML), used to markup Internet applications for handheld devices like mobile phones, is written in XML.

Self-Check Exercise

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

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

3) Discuss HTTP protocol for information communication over WWW.

.....

.....

.....

.....

14.5 WEB 2.0

14.5.1 Definition and Need

The term Web 2.0 is given by Dale Dougherty, Head of Maker Media division of O'Reilly in 2003. Since then it has become a popular concept. Web 2.0 refers to the second generation of the Web, which enables people with no specialised technical knowledge to create their own websites, to self-publish, create and upload audio and video files, share photos and information and complete a variety of other tasks.

While there is no set definition of Web 2.0, it generally refers to the use of the web more as a social platform where users can participate by generating their own content alongside the content provided by the websites.

"Web 1.0 was all about connecting people. It was an interactive space, and I think Web 2.0 is, of course, a piece of jargon, nobody even knows what it means" (12). The above two definitions and excerpt from the talk with Tim Berners Lee show that there is no hard and fast definition of Web 2.0 however it is agreed that it is the second generation of Web where users can participate in generation of web content with much of the knowledge technology. The spectrum of web based content and services cover online shopping, email, chatting, discussion forums, blogs, wikis, social networks, YouTube, Twitter and so on. The products are more personalised now compared to more generic as it used to be.

Another aspect of Web 2.0 is the generation of product and services which are specialised and more users' centric. The trend is to develop tailor made products on demand like online e-learning modules; information digests systems, literature review etc. that are more user oriented services. The Web 2.0 has empowered users to be more interactive with the existing services over World Wide Web.

14.5.2 Need and Features

Following points are instrumental towards the development of Web 2.0 technologies:

- Users' Participation;
- User centric Services;
- Decentralisation and Interoperability;
- Hiding Technological Complexity;
- Modularity.

Users' Participation

With legacy web, the communication was more from source to users where as reverse was not possible. Web 2.0 provides means and tools which empowers users not just to communicate back to the source but also generate content for the website. A user can express their feelings and view about the content. They can agree or disagree with the source. This participation leads in collaboration and development of innovative thoughts.

User Centric Services

The trend of services and its delivery is transformed in Web 2.0 environment. Earlier it was technology, which used to ride the market. But with the advent of new technologies, web services are more customer/user oriented. These services require enough flexibility in their modus-operandi to meet the ever changing needs of users. The user may keep changing their priorities and services options and the services have to stand themselves to each call of their users.

Federated and Interoperability

The applications in Web 2.0 environment are distributed over different nodes. And each node is responsible for its own services. However, collaboratively all the nodes can generate one service through a single platform. The service nodes generate services in a standard format, which can be amalgamated to a single service. This makes systems interoperable over a federated environment.

Hiding Technological Complexity

Web 2.0 platform aid their users to create services or contents without knowing much

of the technology. Thus, it is more user-centric. Web 2.0, in principle, is really not about the technological complexity though technology plays a major part from backend. The technology is developed in such a way that users need not bother about the hassles of technology rather they should concentrate on content of the services. The technology is kept hidden from the user.

Modularity

Modules are components of any system. A system functions in coordination of different modules to offer a service or product. Modular approach towards a system provides flexibility for adding or removing any feature out of the system. Legacy Web was about providing information in a robust but inflexible way. In Web 2.0, the modularity provides facility to add or remove components offering flexibility to a great extent.

Features of Web 2.0 Applications

Tim O'Reilly suggests that the true test of a Web 2.0 service relies on amalgamation of some or all of the following features: (18)

- services, not packaged software, with cost-effective scalability;
- control over unique, hard-to-recreate data sources that get richer as more people use them;
- trusting users as co-developers;
- harnessing collective intelligence;
- leveraging the long tail through customer self-service;
- software above the level of a single device; and
- Light-weight user interfaces, development models, and business models.

Self-Check Exercise

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

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

4) Write features of Web 2.0 applications.

.....

.....

.....

.....

14.5.3 Web 2.0 Applications

Collaborative Web

One of the important features of Web 2.0 is Collaborative services. In a collaborative approach content is generated by a number of persons working at different locations. Online Wikipedia is a good example of collaborative content development.

- **Wikipedia**

Wikipedia (23) is a community-based encyclopedia where anonymous volunteers from Internet contribute articles. It is free to use and one can write as well as change the

Wikipedia articles. The project was started in 2001 and has become one of the most consulted website over Internet.

It is a live service, which is being continuously updated. Newer topics are being added every moment. However the control over the content is kept by editing process to stop any kind of vandalism and misinformation.

The topics in the Wikipedia is arranged in classified manner with following broad subjects:

- Humanities
- Social Sciences
- Natural Sciences
- Formal Sciences
- Professions and Applied Sciences

Each subject is further divided into narrower subjects. Articles are the leaf nodes, which are interlinked as and when, referred within the text.

Each article starts with a small introduction followed by a content of topics covered in the article. At the end, list of references, suggested readings (See also), External links (links outside Wikipedia website) and Further Readings are given to facilitate the reading.

Example: http://en.wikipedia.org (Fig. 14.6)

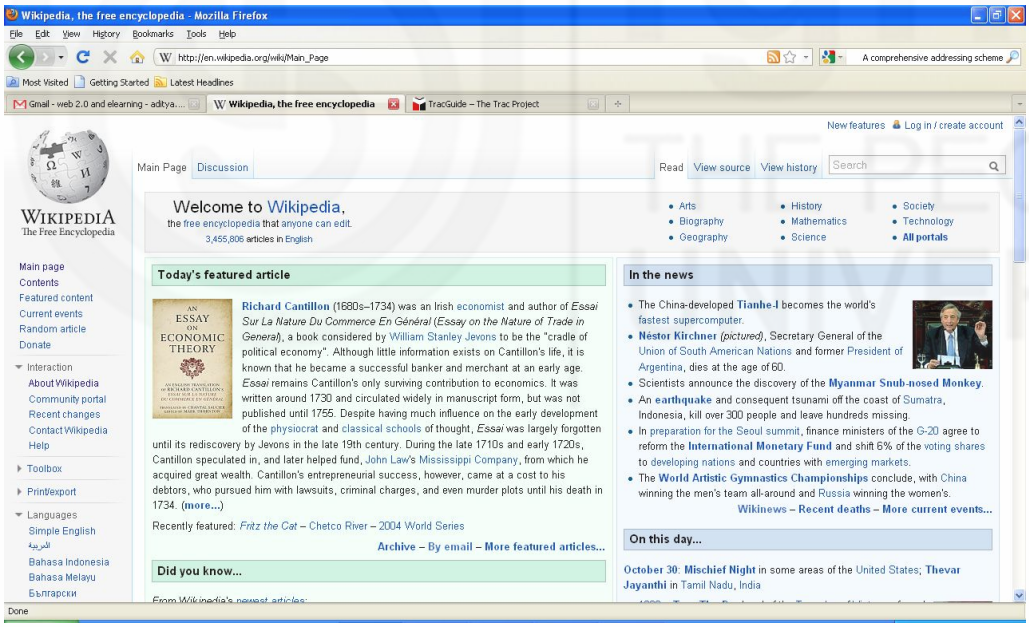

Fig. 14.6: Wikipedia

- **Blogging**

Phenomenon of blogging has picked up immensely among new as well as older generation. According to a statistics, a new blog is created on average every second; 54,000 posts are created every hour, translating into roughly 1.3 million new posts per day. The area or the usage where blog is created is known as blogosphere. A blog promotes freedom of speech, interaction with audiences, and a tool for aggressive marketing.

Blogger offers service to host individual blogs online. Initially, it was started as a small company but now owned by GOOGLE. It facilitates people to express their thoughts online.

Another important blogging website is Wordpress.com (Fig. 14.7) which allows hosting of individual blogs and maintains it.

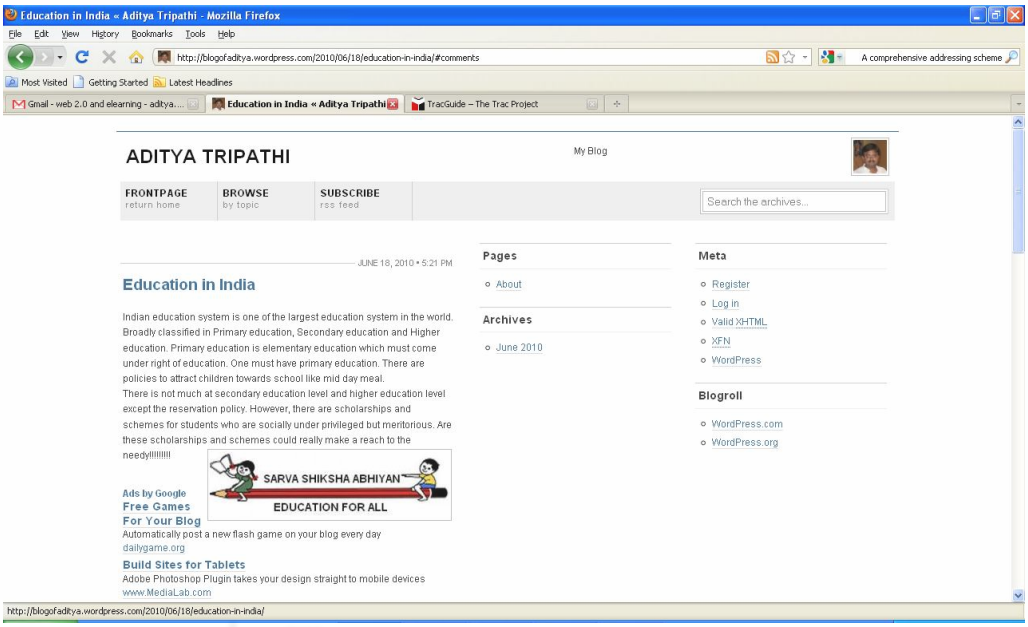

Fig. 14.7: An Example of a Blog at Wordpress.com

These blogging websites provide an administrative panel to administer one's own blogging site. Normally, two kinds of subscription are offered by these blogging sites, free and fee based subscription. The later one is used by commercial organisations, celebrities and so on whereas free ones are available for all. The topics in a blog can be classified and comments can be moderated by the blog administrator.

Example of Blogs: http://www.blogger.com/
http://wordpress.com/

- **Project Management System**

Managing a project requires collaborating different activities together with definite time schedules. Particularly, during the project different versions of products are developed which has to be tracked so that they do not create confusion.

A project management system is ideal for managing project of software development. However, it can manage projects apart from software development. These systems support charting down the work-plan and share it with all the members. These systems use version control for a particular set of file or software and highlight the differences among the different version. These systems have strong user control mechanism, which does not allow a non-authenticated user to look into the areas, which he/she is not concerned with. These systems highlight the landmarks and project the time frame for the stakeholder participating in the project.

Trac is an open source lightweight project management tool, which can be implemented as a web based API. Trac supports to monitor and resolve individual bugs, issues, feature requests, and ideas. It has a ticketing system, which is numbering system for any of the issues mentioned during the project. Trac has supervision module with built-in documentation server, which can be used to keep of documentation in a form of a Wiki.

Bugzilla is another Defect tracking system, used to track the bugs of software. It has a feature to notify through email about the changes in the versions of software system.

Interactive Web

Interactive tools of Web 2.0 are used for collaboration of users to share their ideas and work. It covers all the tools, which are used for sharing works and conducting online discussions and meetings.

- **Online Chat**

There are many free applications, which can be used for online chatting using text, audio and video. Important chatting applications are, Skype, GTalk (or GoogleTalk), Yahoo Chat. These applications are free to use over computer-to-computer chatting. One can hold conferencing as well as one to one chat. These applications can be also used to transfer the files over Internet.

- **Document Sharing Tool**

These are tools which can be used to share documents, images and audio visuals over Internet.

Document Sharing: GoogleDoc

It is free tool for sharing documents over Internet. It supports online upload, editing and sharing of documents, spreadsheet, presentation, drawing and forms. It supports the following functionalities:

- Real-time collaborative highlighting in documents
- In-cell dropdown and cell validation in spread-sheets
- Shows all formula
- Spell checker
- Page sizes

- Auto-linking the text in the documents
- A new curve tool in drawings
- Convert files in the document list
- In built dictionary – one can add terms in the dictionary also
- Create Tables and draggable rows and columns to resize
- Document translation
- Searching of text in the document
- Mobile support
- Format painter in spreadsheets
- In built Optical Character Recognition (OCR) for PDF (Portable Document Format)
- A keyboard shortcut pop-up and more in drawings
- Including drawings in the text
- Sorting of text
- Create forms

GoolgleDocs (Fig. 14.8) has almost all the functionalities to support document texting and sharing with a group of people.

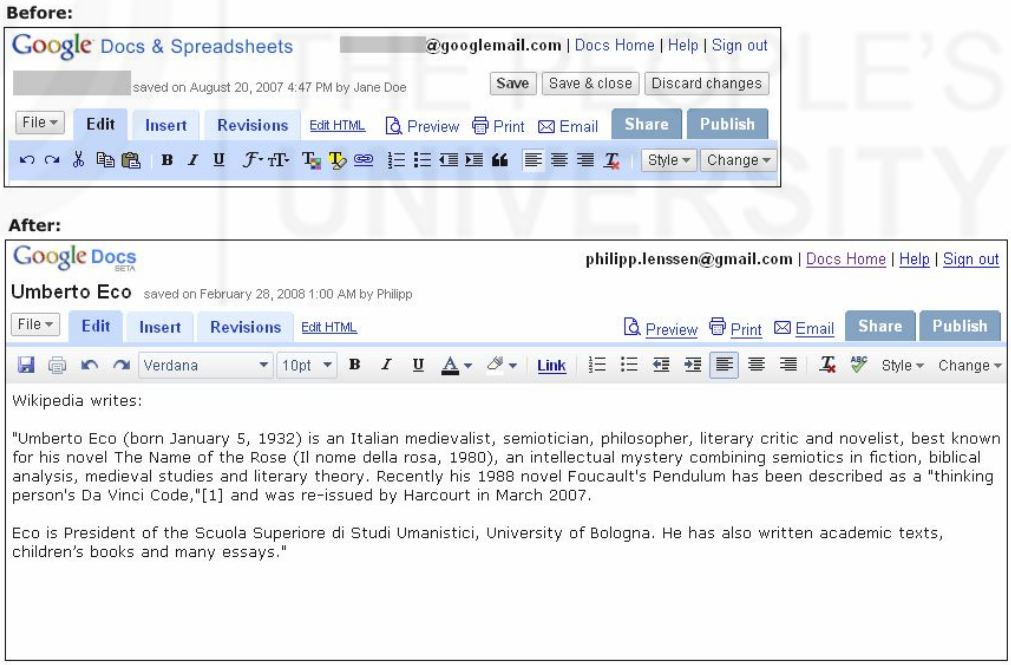

Fig. 14.8: GoogleDocs

Image Sharing: Picasa

Image sharing is an important feature of Web 2.0. People can share their albums online, which can be viewed online by others. One can define the level of access to their photographs. Local album on computer can be synchronised with online album using a client software and any addition in the local album will be uploaded to the website. Picasa (Fig. 14.9) is an online image sharing tool by Google. Another example of image sharing is Flicker.

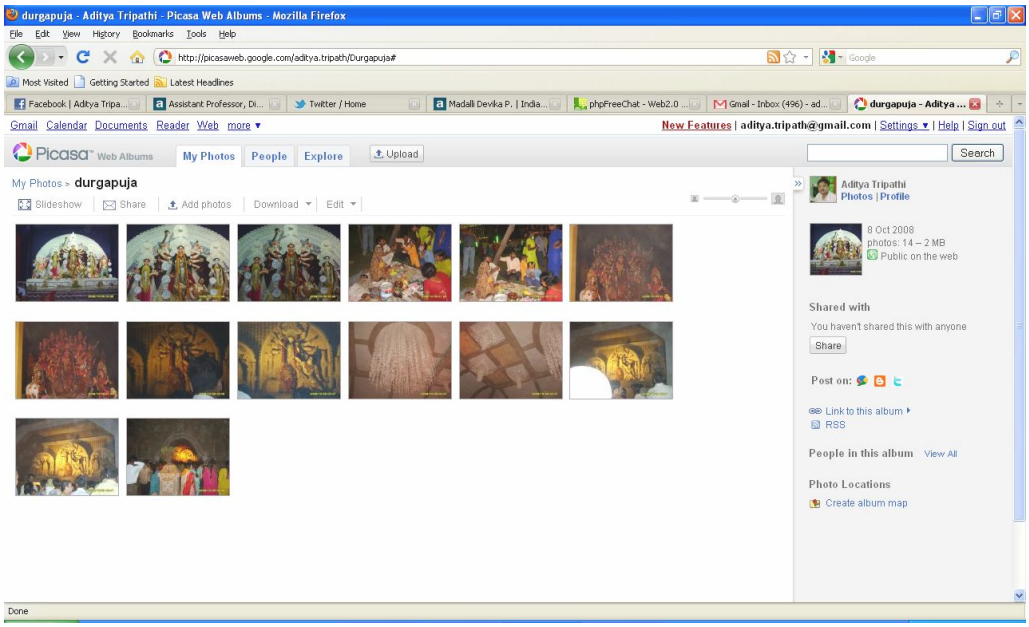

Fig. 14.9: Picasa

Movie Sharing: YouTube

It a website to share video clips (Fig. 14.10). Presently, it is owned by Google and acting as subsidiary firm to Google. It was developed in 2005 by some of the employees of PayPal, a website for making online transaction. This technology uses Adobe Flash Video technology to display a wide variety of user-generated video content, including movie clips, TV clips, and music videos, as well as amateur content such as video blogging and short original videos. Individuals have uploaded most of the content on YouTube. Many of the media corporate also do offer their content through YouTube. Anyone can watch the video, as they are available in open domain free of cost.

Fig.14.10: YouTube

Social Networking

Social networking is a remarkable phenomenon of Web 2.0 implementation. People can connect online with each other. They can search for old friends or form a group of

people with same kind of interest. Such sites are not only used by common man to connect to mass or by celebrities but they have become a tool for advertising firms and products. These social networking sites contain several applications like, image sharing, video sharing, online chatting, emailing, commenting or sending public as well as private messages to an individual. These sites are available over mobile applications in mobile phones so that one can keep connected to people even when he/she is travelling.

- **Facebook**

With all the above mentioned features, Facebook (Fig. 14.11) presently hosts more than 500 million active users and on any given day 50% of the registered users use Facebook. It has become such an important phenomenon for the people using Internet that on average each user has 130 friends or people to which they are connected with. One can share any object with others, which includes pages, groups, events and community pages. It has become a largest sharing platform where around 550 thousand objects are being shared globally. The popularity of Facebook is so much that it is being translated by user community voluntarily using its translation application. Facebook is widely available as an application in mobile phones of various brands. There are more than 200 mobile operators in 60 countries working to deploy and promote Facebook.

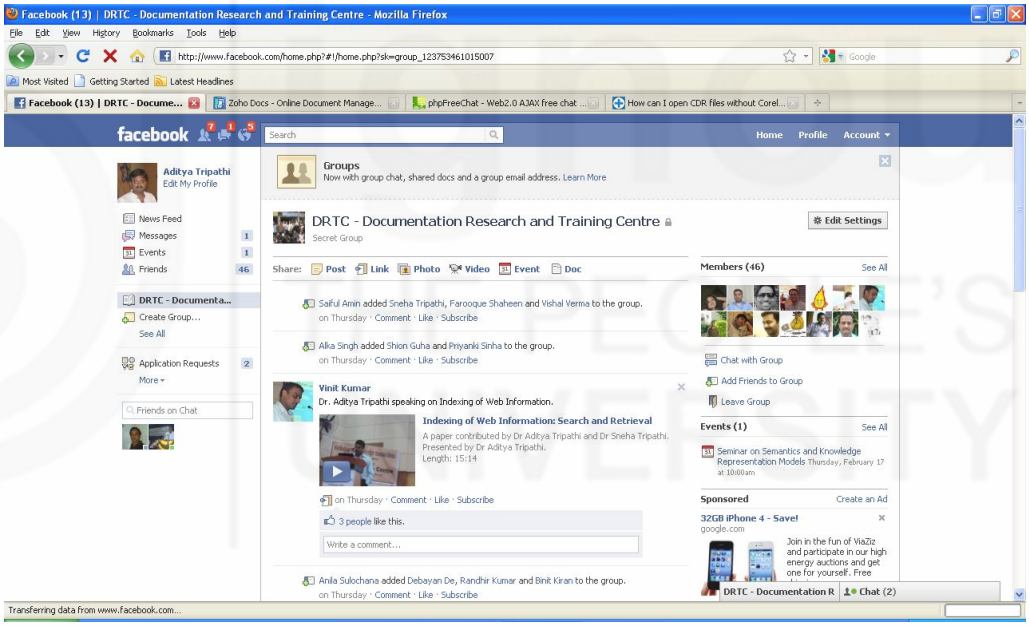

Fig. 14.11: A Facebook Account

Twitter and Orkut are some of the other examples of social networking website. Twitter is very popular among the celebrities. People can follow each other over Twitter.

- **Academia.edu**

This is a social networking site specially designed for the people with academic interests. They can share their ideas with others. It facilitates sharing research interests by hosting published works of their users. He/she can share research papers and articles. It also sends notification to its users if others are searching them. It collects jobs and other academic opportunities for the user as per his/her interests. It is very popular socialising site among the academic and research community.

Information Mashup

These are website which uses the content and services from different service provider about different aspects of subject. For example, if someone is looking for Delhi. A

Mashup website would provide information about the map of Delhi, Weather Forecast, News from Delhi, Photographs of important places, mode of transport, administrative set up and so on. The number of the services, which can be included in Information Mashup, is never ending. Mashup is important to make already existing data more useful, pertinently for personal and professional use. These are hybrid services and often look as an advanced version of web portals, which used to host static webpages.

- **Indianrailinfo**

Indianrailinfo (Fig. 14.12) is a mashup of different kind of information about Indian railways. This website collects information from different websites of Indian railways. It is one stop portal for information about arrival/departure of trains from a station, PNR status, seat availability and so on. IndianrailInfo is also connected through Google, Yahoo and Facebook and it can extract user's information through these websites.

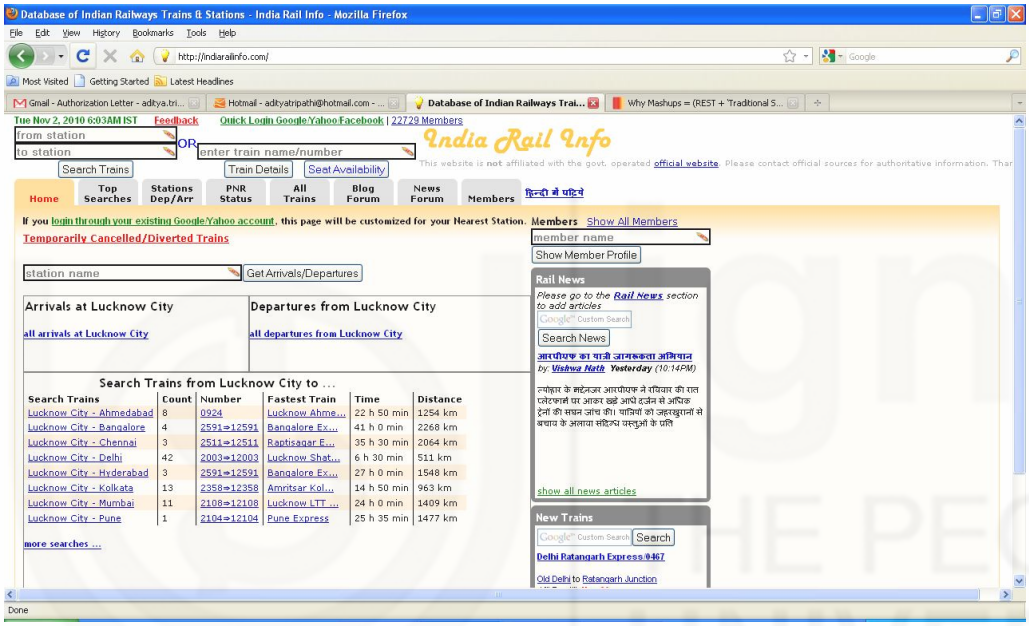

Fig. 14.12: Indianrailinfo

Self-Check Exercise

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

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

5) Name different types of Web 2.0 applications.

.....

.....

.....

.....

14.6 IMPACT OF WEB 2.0 TOOLS OVER WWW AND SEMANTIC WEB

Though there is no clear demarcation between Web 1.0 and Web 2.0 but the product and services offered under the umbrella of later is quite different from the previous. The services are more focused and interoperable in nature. One application or web service

could exploit the data from other. These services are wise enough to correlate the data of users need and services offered by service providers. Semantic web is about developing such kind of services or agents, which are more meaningful from the users' point of view. Implementation of web ontology can bring in more contexts in the pursuit of information and web-based services from the user's point of view.

14.7 SUMMARY

In the present chapter we learnt how the present Web has evolved from a simple idea of Sir Tim Berners Lee. It has become a platform of many complex services like, blogs, mashups, and social networking websites. The existing technology is still in evolutionary stage and more is expected to come in the form of Semantic Web. In the present chapter we could create understanding about,

- The framework on which Web functions
- The languages which are used to develop web applications
- Various Web 2.0 based services.

14.8 ANSWERS TO SELF CHECK EXERCISES

1) Web Browser is an application software to view webpages over Internet. It is used to retrieve, view and transfer information over WWW.

2) Addressing protocols is a standard to identify a web object by its names and addresses. It is known as Uniform Resource Identifiers (URIs). Uniform Resource Locator (URL) is a kind of URI which is used for locating web objects or webpages over WWW. It is a global addressing system, which is translated to IP address in order to locate the server containing information. The first part of the address is known as protocol identifier and it indicates what protocol to use; the second part is called as resource name and it specifies the IP address through the domain name where the resource is located. The protocol identifier and the resource name are separated by a colon and two forward slashes.

3) HTTP provides mechanism to address a website. It is a networking protocol for distributed, collaborative, hypermedia information systems. HTTP is used for data communication over World Wide Web. It supports retrieval, which includes search, front-end update, and annotation. HTTP protocol is based on a request from a client and response from a server. A server hosts webpages and on request delivers it to the client. This kind of request/response format of communication is known as Client – Server Architecture. Every request is treated as a new request and there is no dedicated connection between client and server. This is known as stateless connection. HTTP includes support to SMTP, NNTP, FTP, Gopher, and WAIS protocols. The last two are no more in use for real world web communication. Hence, HTTP allows basic hypermedia access to resources available from diverse applications.

4) Following are the key features of Web 2.0 applications,
   - Users' Participation
   - User centric services
   - Decentralisation and Interoperability
   - Hiding technological complexity
   - Modularity

5) Different types of Web 2.0 applications are:

- Collaborative Web Tools – Wikipedia, Drupal, Joomla, Blog, Trac, Bugzilla etc.
- Interactive Web Tools – Gtalk, Yahoo Chat, GoogleDoc, Picasa, Flickr
- Social Networking Tools- Orkut, Facebook, Twitter etc.
- Information Mashup – Google Map, Indiarailinfo, etc.

14.9 KEYWORDS

Blog : A website where entries are displayed in a reverse chronological order. A Blog is used to communicate to the audience and reverse. A blog may use text, images, and links to other blogs, web pages, and other media related.

eLearning : Stands of Electronic learning. E-learning is a method to impart distance education through electronic media which includes, online and offline media.

Flickr : It is a photo sharing website and web services where members can share photographs. The photographs uploaded on the website can be tagged for efficient retrieval.

Folksonomy : Folksonomy is community based labeling or tagging system performed by the Internet community (users). It is an open-ended labeling system for web objects including web pages, photographs, software and so on. This tagging system is intended to make a body of information increasingly easier to search, discover, and navigate over time.

Gopher : A distributed document search and retrieval network protocol designed for the Internet, which is now obsolete.

Intelligent Agents : A piece of software, which has intelligence to draw the inference from the fact.

Learning Management System (LMS) : Piece of software that enables the management and delivery of learning content and resources to students. Normally, these software works in online environment.

Listserv : An email-based mailing list.

Podcasting : Multimedia file distributed over the Internet using syndication feeds (RSS), for playback on mobile devices and personal computers.

RSS : Stands for Really Simple Syndicate. A standard mechanism of offering content, which can be read

by web browsers as bookmark. It is a simple XML-based system that allows users to subscribe to their favourite websites.

Semantic Web : An idea of more meaningful web where resource discovery will be assisted by intelligent agents.

YouTube : A Web 2.0 based service to share and view video clips.

Web 2.0 : New generation of Web where users can have better control and voice towards using web services and products. The services are more user centric and interactive.

Wiki : An effective tool for collaborative authoring.

14.10 REFERENCES AND FURTHER READING

1. Arms, William. *Digital Libraries*, M.I.T. Press, 1999. <http://www.cs.cornell.edu/wya/DigLib/MS1999/Chapter2.html>
2. Bugzilla. < http://www.bugzilla.org>
3. E-mail. (n.d.). Available at *Wikipedia*: <http://en.wikipedia.org/wiki/E-mail>
4. Enhancing Web 2.0 literacy: Using blogging as a management tool. *Strategic Direction*. 24.11 (2008) pp. 12-13. DOI: 10.1108/02580540810914977
5. Extensible Markup Language (XML) 1.0 Available *at World Wide Web Consortium* (1998) < http://www.w3.org/TR/1998/REC-xml-19980210.html#sec-intro>
6. Flynn, Peter (ed.). *The XML FAQ: Frequently-Asked Questions about the Extensible Markup Language*. (2010) < http://xml.silmaril.ie/basics/markup/ >
7. Glossary. Available at *European Union's Publisher Website* <http://publications.europa.eu/vademecum/vademecum/9313fdfe-c49e-119e-45c6a6441e63e066_en.html>
8. Gralla, P.. *How the Internet works*. 4th ed. Indianapolis: Que Publishing, 1998. Print
9. How the web began. (n.d.) Available at CERN website. <http://public.web.cern.ch/public/en/about/webstory-en.html>
10. HTTP Properties (2006) <http://www.wdvl.com/Internet/Protocols/HTTP/Properties.html>
11. Johnson, Tony. *Spinning the World-Wide Web*. (1994) <http://www-sld.slac.stanford.edu/sldwww/beamline/article.html#concepts>
12. Laningham, Scott (ed.) IBM developerworks Interviews: Tim Berners-Lee. (2006). <www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html>
13. Learn The Net: E-mail. (n.d.). <http://www.learnthenet.com/english/section/email.html>

14. Levine, J. R., Young, M. L. & Baroudi, C. *The Internet for Dummies*. 11th ed. Hoboken: Wiley Publishing, 2007. Print

15. Lindsell-Roberts, S. *135 tips on email and instant messages: plus blogs, chatrooms, and texting*. Boston: Houghton Mifflin Company, 2008. Print

16. Multipurpose Internet Mail Extensions. (n.d.). Available at *Wikipedia.* <http://en.wikipedia.org/wiki/Multipurpose_Internet_Mail_Extensions>

17. Nations, Daniel. *What is Web 2.0? How Web 2.0 Is Defining Society.* (2010). <http://webtrends.about.com/od/web20/a/what-is-web20.htm>

18. O'Reilly, T. *What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*. (2005). <www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>

19. Simple Mail Transfer Protocol. (n.d.). Available at *Wikipedia*. <http://en.wikipedia.org/wiki/Simple_Mail_Transfer_Protocol>

20. The history of email. (n.d.). Available at *Net History*. <http://www.nethistory.info/History%20of%20the%20Internet/email.html>

21. Thomas, Brian J. *The Internet for scientists and engineers: online tools and resources*. 3rd ed. Oxford: Oxford University Press, 1997. Print

22. Trac. < http://trac.edgewall.org/ >

23. Wikipedia< http://www.wikipedia.org/>

24. Wood, D. *Programming Internet Email*. Sebastopol: O'Reilly & Associates, 1999. Print

25. Young, Linda. A social media glossary. (2008) < http://www.capilanou.ca/help/login-page/active-cms/glossary.html>

26. Yves, Lafon. *HTTP - Hypertext Transfer Protocol Available at World Wide Web Consortium*. (2010) < http://www.w3.org/Protocols/Overview.html>

UNIT 15 SEARCH ENGINES

Structure

15.0 OBJECTIVES

After going through this Unit, you will be able to:

- describe different search tools available to locate information on web;
- discuss search engines as an effective information retrieval tool;
- know different search strategies to increase the performance of search tools; and
- explain search engine architecture.

15.1 INTRODUCTION

Internet has created revolutionary changes in this era of Information Technology. For many, it is one stop platform to find or locate any information they are interested in. Traditionally, librarians had the job to assist their users to locate the information they needed. But, now the scenario has changed a lot. Internet has in offer a variety of search tools such as search engines, search directories to locate the information on web.

A search on web is a simple process and can be conducted by simply issuing a query to the search tool. The search tool in return will look for the information in its web based information databases and retrieves those, which are relevant to the query. Searching is an iterative process i.e. one needs to keep working on their query unless the exact information is located.

The very first tool used for searching on the Internet was Archie. The name stands for "archive" without the "v." It was created in 1990 by Alan Emtage, a student at McGill University in Montreal. Veronica (*V*ery *E*asy *R*odent-*O*riented *N*et-wide *I*ndex to *C*omputerised *A*rchives) and Jughead (*J*onsy's *U*niversal *G*opher *H*ierarchy *E*xcavation *A*nd *D*isplay) were two other popular search programs.

There are three basic types of search tools that most people use to find what they are looking for on the Web: Search Engines, Subject Directories and Meta Search Tools. Search Engines are more generic and much larger than Subject Directories. Meta Search Tools get their results from several search engines. The following sections will provide an elaboration on these search tools.

15.2 SEARCH ENGINES

Search engine is a tool for locating information from a collection. Search engines uses information about the information (such as metadata, catalogue) stored in the database to locate information. Sometimes they perform full text search within the document from first character to last character.

The search is done on pattern matching algorithm whether it is a database or full text.

15.3 TYPES OF SEARCH TOOLS

15.3.1 Search Directory

Search directories are classified collections of documents. They are good for searching with a context. These directories are good for browsing. In subject directories, documents are pre classified by a person. Librarians' Internet Index; Google Directory; Yahoo!; dmoz are some of the examples of subject directories.

There are two basic types of directories:

- Academic and Professional Directories: These are often created and maintained by subject experts to support the needs of researchers. INFOMINE, from the University of California, is a good example of an academic directory.
- Commercial Directories: These cater to the needs of general public. Directories of Yahoo! and Google are examples of commercial directories.

Fig. 15.1: Librarians' Internet Index (www.lii.org)

15.3.2 Search Engines

World Wide Web is a network of several information databases. In recent years, an exponential growth in these databases has made it difficult to locate a particular piece of information. Internet offers a powerful tool known as search engine to manage, filter and retrieve the information for their users.

Search engines are automated tools for searching information from a collection using metadata stored in the database of search engine. In other words, it is an information retrieval system and assists in locating information on web.

Google and Yahoo! are most popular search engines.

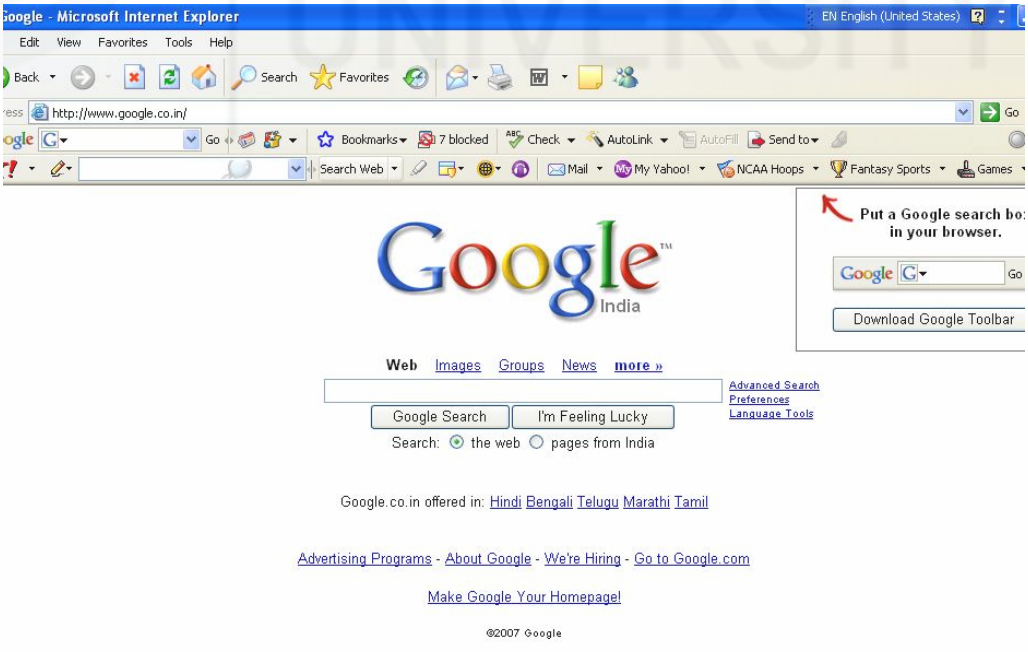

Fig. 15.2: Search Interface of Google

15.3.3 Meta-search Engines

Meta Search engines are online tools (search engines) which performs simultaneous search on more than one search engine at a time. These search engines aggregates the results into a single list and displays them according to their source. e.g. Dogpile is a metasearch engine and gets its results from Google, Yahoo, MSN Search, Ask , About, MIVA, LookSmart, and more.

Example: Dogpile, WebCrawler, Browsys

Fig. 15.3: Search Interface of "Dogpile", a Metasearch Engine

Self-Check Exercise

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

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

1) Write a short note on Search Tools?

2) Name any four search engines and two metasearch engines?

.....

.....

.....

.....

15.4 FEATURES OF SEARCH TOOLS

The searching tools follows pattern matching algorithm. There are many types of searches can be done using search tools,

15.4.1 Keyword Search

When searching is done using a keyword it is known as Keyword searching. Keyword may occur at any place in the document or in the metadata field. This kind of search has higher recall value.

15.4.2 Boolean Search

Logical AND, OR and NOT are known as Boolean operators. When Boolean operators are used for searching it is known as Boolean search. The operators are used for combining more than one word with certain conditions. These kind of searching also known as Combinatorial search.

- **AND**

This operator will retrieve all the documents which contains all the keywords occurring at both ends of the AND operator.

Syntax: <Search Term **A**> AND <Search Term **B**>

Example: Library ***AND*** Information

Output:

1) The above query will retrieve only those documents which contains both the terms Library and Documentation
2) The precision in search is more. The number of documents retrieved will be less hence less is the recall value.

- **OR**

This operator will retrieve all the documents which contains all the keywords occurring at both ends of the OR operator.

Syntax: <Search Term **A**> OR <Search Term **B**>

Example: Library ***OR*** Information

Output:

1) The above query will retrieve all documents which contains both the terms Library and Documentation
2) The recall in search is more. The number of documents retrieved will be more but the precision in retrieved documents will be less.

- **NOT or AND NOT**

These operators increase the precision of the search result. The query can be made more specific by using these operators. Using the capitalised AND NOT operator preceding a search term eliminates documents that contain that term.

Syntax

<Words to be searched> AND NOT <Words not to be searched>

Example:

If user is looking for information on Drivers and do not want documents that include information relating to the Screw Drivers the query could be "Driver" AND NOT Screw.

15.4.3 Proximity Search

This is another kind of Combinatorial search where the proximity of two words is checked. The term proximity means 'nearness of words'. Proximity is given in terms of number of words by which two words should be separated. There are two kinds of proximities,

1) Near Proximity

2) Exact Proximity

1) **Near Proximity**

Near Proximity brings range of search results where the number of proximity is from adjacent to the mentioned number proximity. For example, for two keywords COLLEGE and LIBRARIANS, if the proximity of 3 is applied between them it will bring the results as follows,

COLLEGE LIBRARIANS

COLLEGE FOR LIBRARIANS

COLLEGE OF LIBRARIANS

COLLEGE WITHOUT LIBRARIANS

COLLEGE WITH THE LIBRARIANS

COLLEGE OF INDIAN LIBRARIANS

COLLEGE OF THE LIBRARIANS

The Near Proximity would bring the results where the search terms would be separated by no word to n-1 word (where n is number of proximity). In other words, near proximity brings all the proximities which are lesser than the mentioned number.

2) **Exact Proximity**

Exact proximity brings the results with exact number of proximity mentioned. It does not bring the results which have lesser number of proximity. For example, if the number of proximity is set to 3 between two keywords i.e. COLLEGE and LIBRARIANS. The retrieved result would be,

COLLEGE WITH THE LIBRARIANS

COLLEGE OF INDIAN LIBRARIANS

COLLEGE OF THE LIBRARIANS

15.4.4 Truncation Search

Truncation means concatenation of words. In other words, if the root string of the words is searched it brings all the derivatives derived out of the given root string. Truncation is of three types based on truncation techniques:

1) **Left Truncation**

When the root string is concatenated from the left side, it is known as left truncation. For example, if the left truncation is implemented for the root string ISM, it will bring all the words which ends with the string ISM, like

BRAHAMINISM

COMMUNISM

SUPHISM

2) **Right Truncation**

When the root string is concatenated from the right side it is known as right truncation. For example, right truncation is used with the root string CLASS, it will bring all the words which starts with the root string CLASS, like

CLASS

CLASSIFICATION

CLASSIFICATIONIST

CLASSIFIER

15.4.5 Case Sensitive Search

One of the major features of search tools is their support to search words based on their case. In other words, search tools can differentiate between Upper and lower cases. For example, DUKE and duke will bring different search results based on the case. In an ordinary/plain search, search tool performs searching irrespective of their cases. However, if case sensitive search is invoked, search tool brings exact search string based on the case of search string.

15.4.6 Limiting Search

There are certain conditions based on which the searches can be narrowed down, for example, by Date, by Domain, by media type, by Document Directory Depth, by Page Depth and so on. This kind of condition reduces the number of search results and increases the relevancy of final output.

15.4.7 Fields Search

Field Search is a kind of **limiting search** to a particular field of the database. Searching can be done within a given context. For example, searching within Title or searching within Author or searching within both the fields. This kind of searching is known as Field search.

15.4.8 File Types Search

When the searching is restricted to a particular file type like, MS-Word, PDF, PPT etc. it is known as file type search.

15.4.9 Stop Words

While searching documents in a collection or over Web, some frequently occurring words like prepositions, conjunction etc. should be avoided. In order to avoid such redundant words from the search results, the tools contain a file called stop word file. This file lists all the words which are to be avoided from being indexed. This saves space of storage and reduces time of search.

15.4.10 Ranking

Search tools present the search results in some order. Normally, when system is small

the presented results are arranged in alphabetical order. But when results run in several pages, it becomes important to present the most relevant document on the top followed by less relevant one. Thus, it is important to rank the retrieved documents based on their relevancy to their users. Search engines have an automated mechanism to rank the retrieved results according to the relevancy of each retrieved search result.

For example, PageRank of Google is an algorithm for measuring weightage of results based on link analysis.

15.4.11 Family Filters

Family filters are used to reduce, if not remove, the objectionable matter to appear on search results. Search engines do provide functionality for setting family filters as safe search. Family filters are used by Google (as Safe Search), Altavista, Yahoo and so on. Apart from the search engines there are tools (e.g. Naomi, which is a freeware) which can be loaded on computers to stop display of obscene matters.

15.4.12 Fuzzy Search

Fussy search is one of the major features of today's information retrieval system. It brings out results based on approximations. In other words, these are error correction algorithms. For example, if a keyword is miss-spelt search algorithm used in searching attempts to render the search result according to correct spelling. Such algorithms are known as Soundex and Metaphone algorithms. Levenshtein distance algorithm is one such kind of algorithm used by Lucene search engine, an open source search tool.

Self-Check Exercise

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

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

3) Write short note on Boolean Search?

4) What is a Proximity Search?

5) What is "PageRank"?

.....

.....

.....

.....

15.5 ARCHITECTURE OF SEARCH TOOLS

WWW is a huge source of information and search engines are tools or agents for locating information. Everything and anything can be located over Internet using search engines. Search Engines are tools which provide a kind of interface for users to search the web. A Search Engine basically has three components:

- Web Crawler
- Metadata storage
- Search Agent

15.5.1 Web Crawler

Web crawler is also known as robot or spider. It is a program which goes to each and every site over Internet and indexes the content of the webpage. The content includes metadata information and the text from the webpage. Text of the page can be indexed as a whole or only few lines or bytes of data are stored. This index is stored within search engines database with corresponding URL (Uniform Resource Locator).

Hence, Web crawler is a program used by search engine in order to extract data from the web pages so that pages can be searched using the search engine's interface. Following are the names of web crawlers used by popular search engines:

Search engine	Robot
Google	Googlebot
Yahoo	Slurp
MSN	MSNbot

15.5.2 Metadata Storage

Metadata is data about data. Web Crawlers extract metadata like title, author, filename, file type, file size, links and so on from search engines. The metadata information is extracted from meta-tag of webpage, file name, file extension, etc. The collected metadata is stored in the repository of search engine, which is a database, in the form of index. Normally, many of the system follow a kind of keyword indexing. But the Keyword indexing is good for recall not for precision. In such cases the context for search is lost. However, the use of metadata is good for preserving the context of search term. The metadata is stored in the form of metadata index inside the search engine database.

Fig. 15.4: Architecture of Search Engine

The most commonly used metadata schema is Dublin Core Metadata Initiative (DCMI) over Internet. The standard is developed and maintained by DCMI and DCMI Task Groups. There are 15 elements given in Dublin Core. Apart from these 15 elements there are other metadata set vocabularies which should be used with 15 elements.

Title	Documentation Research & Training Centre – Home page
Creator	Aditya Tripathi
Publisher	Documentation Research & Training Centre
Identifier	http://drtc.isibang.ac.in
Subject	Library and Information Science
Format	txt/html
Language	English
Rights	Indian Statistical Institute

Fig. 15.5: Dublin Core Metadata Record

15.5.3 Search Agent

Search Agent is a set of search programs which receives query from the user interface. The received key word is passed through several algorithms. Some commonly used algorithms are as follows:

- Boolean operator
- Data clustering algorithms
- Error correction algorithms.

Once the query or search term is processed, the agent performs search within the repository of the search engine and retrieves the search result and send to user interface.

Boolean Operators

The Boolean operators are AND, OR and NOT. These operators are used to generate combinatorial search. AND and NOT operators increase precision where as OR increases recall of search results. The shaded area represents retrieved records in the following example (Fig. 15.6). Almost all the search engines provide facility of using Boolean operators. These operators can also be used to combine keywords present in different fields.

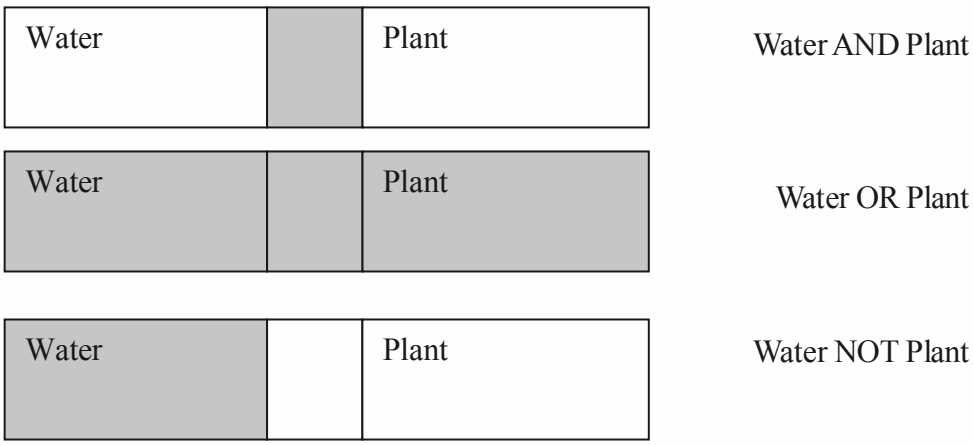

Fig. 15.6: Boolean Operators

Clustering Algorithms

Cluster is group of like entities. Search engines use clustering technique to classify like concepts for more meaningful retrieval. Clustering is a technique by which similar kind

of data are grouped based on certain characteristics. This technique is very useful when there is large set of data for effective and efficient retrieval. For example, the faculty member of a university can be clustered according to their area of specialisation.

The clustering algorithm attempts to identify groups in a given set of data or population based on likeness of certain characteristics or traits. Thus it creates a picture of a big group and then inside several sub groups. The algorithms attempts to identify core entity which is also known as Centroid. Hence, centriod is the centre of the concept or the core concept. The other concepts are placed around the core concept. The likeness decreases with the increasing distance of the concept from the core concept or the centroid. Hence, to determine cluster membership and sise, the algorithm evaluates the distance between concept and the core. Such techniques are highly dependent on use of statistics for generating clusters.

Error correction Algorithms

While typing the keywords often the searcher commits mistake. Search engines are equipped with the algorithms which correct the spelling and yield meaningful results. Google's "Did you mean" uses these kind of algorithms. In case of incorrect spelling or if a word is not found in its database, Google suggests alternate spellings for the keyword.

Soundex and metaphone algorithms are examples of such kind of algorithms. Both algorithms are based on the pronunciation of a word.

In **soundex algorithm** a numeric code is assigned to each character used in a word and when search is performed, words with similar codes are also brought out in search result. Metaphone also works on same algorithm but unlike soundex which encodes a word on letter-by-letter basis, it encodes groups of letters i.e. a word.

Metaphone algorithm embodies more accurately the rules of pronunciation in language. Such algorithms are well established for English as a language. Both algorithms return all words that exactly match the desired word as well as all similar sounding words (homophones).

15.5.4 User Interface

User Interface is the part which interacts with the user. It is like a switch board for the user for invoking the system to perform search for needed information. There are two parts of User Interface:

- Search Interface
- Result Interface

Search Interface

This is the end from where users enter his/her search terms. It is one of the major components which initiate the communication between users and the system. The Search Interface performs following tasks:

1) *Capturing user input/query*

It is also known as front-end. The interface captures keywords given by users. It passes it to the search agent. The look and feel of the front-end should be easy to operate.

2) *Search refinement*

Search interface should have facility to refine the search. The refinement facility should

be given within the displayed search results. Hence, user interface should provide facilities for modifying search statement. Sometimes user interface gives facility for browsing pre-classified categories.

3) *Advance Search statement*

User interface should have another interface for advance level search. The advance level search includes, use of Boolean operators, image search, file-type search, language search, date-wise search and so on.

Result Interface

Display of search results is another important aspect of searching. It should be in user friendly format and customisable by the user.

Ranking of search result is an important feature. The search engines use algorithms for ranking. Google uses an algorithm called PageRank. In this, link analysis is done for ranking of retrieved documents. Some search engines rank their results based on the frequency of occurrence of search terms. Statistical techniques are widely used for this purpose.

Self-Check Exercise

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

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

6) What is a "Web Crawler"?

7) What is Clustering?

8) What is Soundex Algorithm?

.....

.....

.....

.....

15.6 CHALLENGES

Searching on Internet has gone a long way. In the era of Web 2.0, search engines are well equipped with several algorithms like soundex, metaphones, page rank, etc. for improved retrieval of search results. One area which is still required to be improved is context based searching. Though clustering techniques are applied for setting the context but these methods are not fools proof. The implementation of Semantic web is ahead which promises search agents with specialised search strategies and features. These agents would not only perform the search, they will also guide the users in taking decision.

The idea of Semantic web is still evolving and it has still long way to go. But there are products available in the market which has started making mark on the Internet arena like, Wikipedia, Social networking sites, blogs and so on.

15.7 SUMMARY

This Unit highlights the features of search tools. Search engines are web based search tools to search the documents or objects over Internet. The Unit has discussed different

types of search engines i.e. search engine, meta-searching engine and search directory. Search engines have two kinds of search interfaces, simple and advance. Advance search facilitates different kinds of searches such as Boolean search, proximity search, truncation search, case sensitive search, fields level search, file types search, stop words, sorting, etc.

The Unit elaborates the architecture of search engines with reference to user interface, search agents and web crawler.

15.8 ANSWERS TO SELF CHECK EXECRICES

1) To locate information on web, search tools play an important role. The very first tool used for searching was Archie.

   Internet offers various types of search tools:

   a) search engine

   b) subject directories

   c) metasearch engines

2) Search Engine:

   a) Google

   b) Yahoo!

   c) Alta-Vista

   d) Vivisimo (vivisimo.org)

   Meta-search Engine:

   e) Dogpile

   f) Browsys

3) When Boolean operators are used in a search it is known as Boolean Search. The Logical operators such as AND, OR, NOT are known as Boolean Operators.

4) Proximity means 'nearness'. Hence, when a search is based on the proximity or nearness between two words is known as Proximity Search.

5) PageRank is an algorithm used by Google for measuring weightage of results based on link analysis to rank the retrieved results.

6) Web crawler is a program used by search engine in order to extract data from the web pages so that pages can be searched using the search engine's interface. It is also known as robot or spider.

7) Clustering is a technique by which similar kind of data are grouped based on certain characteristics. This technique is very useful when there is large set of data for effective and efficient retrieval.

8) Soundex is an error-correction phonetic algorithm. If there is minor error in the spelling of the search term or there are cases of homophones, this algorithm helps in retrieving the results.

15.9 KEYWORDS

Directories	:	Lists of pages classified into useful categories, (like Yahoo or Looksmart).
Exact Match	:	If a document contains exact match to the query, then only it will get retrieved. Increases precision of result but low recall value.
Query	:	A query is the combination of the word or words used for searching.
Recall	:	Total retrieved records against a query.
Precision	:	Total relevant records retrieved against a query.
Relevance	:	The extent to which retrieved records against a query satisfies the end-user.
Search engine	:	A program that indexes web documents and facilitates user to perform search on them.
Weighting	:	Weighting is a heuristic technique designed to improve the relevance ranking algorithms.
Index term	:	It is a pre-defined term which can be used to refer to the content of a document.
Full Text Search	:	It is a methodology in which all the words which compose the text of the document are used as indexing terms.
Fussy Model	:	It is "set theoretic model" of document retrieval based on fussy theory.
Inverted file	:	An index composed of vocabulary and list of occurrences.

15.10 REFERENCES AND FURTHER READING

1. Besroukov, Nikolai. Searching Algorithms (2009). <http://www.softpanorama.org/Algorithms/searching.shtml>
2. Find Information: Advanced Web Searching. <www.learnthenet.com/english/html/77advanc.htm>
3. Franklin, Curt. How Internet Search Engines Work <http://www.howstuffworks.com/search-engine1.htm>
4. History of Search Engines: From 1945 to Google 2007. <http://www.searchenginehistory.com/>
5. Houghton, Bernard and Convey, John. *On-line information retrieval systems: an introductory manual to principles and practice*. Clive Bingley: London. Print
6. Information Skills Online (2006). <http://infoskills.mq.edu.au/scd/end.htm>
7. Meta Search Engines. <www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html>

8. Subject Directories (2004). <http://www.learnwebskills.com/search/subject.html>

9. Sullivan, Danny. Where Are They Now? Search Engines We've Known & Loved. *Search Engine Watch*, Mar 4, 2003.< http://searchenginewatch.com/2175241>

10. Taylor, Chris. Introduction to metadata (2007). <http://www.library.uq.edu.au/iad/ctmeta4.html>

11. Walt Howe. A brief history of the Internet. <http://www.walthowe.com/navnet/history.html>

UNIT 16 INTERACTIVE AND DISTRIBUTIVE SERVICES

Structure

16.0 OBJECTIVES

After reading this Unit, you will be able to:

- describe interactive services and the benefits associated with them;
- explain prominent distributive services with examples; and
- discuss issues involved with the implementation of these services.

16.1 INTRODUCTION

Interactivity is defined as one to one communication between the two systems. This is the state of a system when the responses actually depend on the inputs received from outside of the system. If we look at the present services available on Web one can easily say that interactivity has been taken care seriously by the developers in the near past. The days of having a system stand-alone and with no interactions are gone. This is the pressure of more involvement of user collaboration causing the implementation of more interactivity. The interactivity further makes the system more user friendly, which subsequently enhances the user satisfaction.

On the other hand, these services provide great chance of collaboration and exchange of information.

16.2 DISTRIBUTED SERVICE

Distributed services are offered through distributed sites. In other words, the machines, which host the services, are placed at different geographical locations. However, these machines interact in a standard manner to offer one or multiple service under one umbrella service. In the subsequent sections we will discuss some of the important distributed services.

16.2.1 Web Directory

The Web contains trillions of terabytes of data and the information is not organised properly. There are some search engines available but sometimes getting answers to questions like, "What resources does the World Wide Web have on Algebra?" or "What kind of information is available on Knowledge Management?" get very tough. The web directory serves as an important service for providing answers to the above-mentioned questions. Web directories are nothing but a topical list of Internet resources arranged in a hierarchical way. Unlike search engines where Web is indexed by using robots and web directories are human maintained and created. Often these people are volunteers or sometimes hired. Generally, web directories are meant to be browsed by subject or topics, they can be searched by keywords too. Some of the popular web directories are listed in table 1 and main page screenshots given in the figures following the table from fig. 16.1 to 16.3.

Table 16.1: Popular Web Directories

Name	Directory URL	Remarks
BUBL Information Service	http://bubl.ac.uk/	From Librarians
Google Web Directory	http://www.google.com/dirhp	
Yahoo Directory	http://dir.yahoo.com/	World's first web directory
DMOS Open Directory Project	http://www.dmos.org/	Public involvement

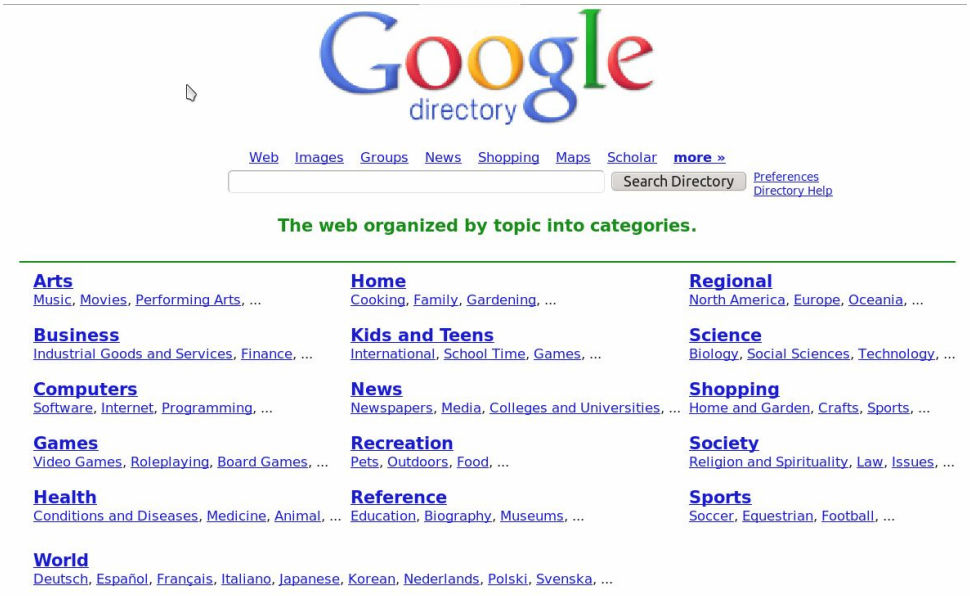

Fig. 16.1: Google Web Directory (http://www.google.com/dirhp)

Fig. 16.2: BUBL Information Service (http://bubl.ac.uk/)

Fig. 16.3: DMOS Open Directory Project (http://www.dmos.org/)

The human element involved in creating and maintaining directories creates both advantages and disadvantages for the user. Some of the advantages are:

i) They contain fewer resources and reduces the information avalanche.

ii) Many directories rate, annotate and categorise the chosen resources.

iii) Directories increase the probability of retrieving the relevant results.

On the other hand, they have some disadvantages too:

i) Most of the directories follow their own hierarchical arrangement, which sometimes leads to arbitrary arrangement too.

ii) Because of human based maintenance the update is not so frequent and sometimes not at all happening.

iii) The subjectivity of the annotation and rating is often under question.

iv) It involves a good understanding or guess about the closest topic about the subject to be researched upon. For example if a user digs in some topic in the directory and finds no resource s/he has to guess or look for another topic.

16.2.2 Bulletin Board

Bulletin or message boards provide a facility for discussion under various topics. They allow individuals to respond to topics or threads in the group, or to begin a new topic or thread by posting a comment or question. The messages posted to a discussion board are permanently visible to everyone who has access to it. Most of the discussion boards implement staged users, the users enter the board as primitive member, in board specific terminology they are known as "newbie" and then by asking questions and posting answers to the question asked by others they get promoted to "starred" members. This is a common phenomenon in many of the Bulletin Boards.

There are millions of bulletin boards available on the Internet. Many news oriented websites, search engines, social networking websites and special interest sites such as people using a particular type of personal computer or sharing an interest or a particular hobby or political issue provide bulletin boards. Some of the major Internet service providers also provide facilities for groups to set up their own bulletin boards and other means of sharing information and communicating among each other. Many libraries provide bulletin boards within their library websites as a means of enabling their users to discuss ideas and share information. Some sites provide bulletin boards for LIS professionals and provide opportunities to share good practices, discuss hot topics or to gain support.

A wide range of software packages are available to enable the use of bulletin boards within websites one such software is PHPBB. Almost all the bulletin boards provide some of the following features:

- They provide a basic search facility, by topic, author and keyword.
- Tools to enable view bulletins in hierarchical format, popularly known as threaded and unthreaded view.
- Facilities to select and save the bulletins.
- Facility to indicate the read and unread messages.

Below are some screenshots of available web bulletin boards:

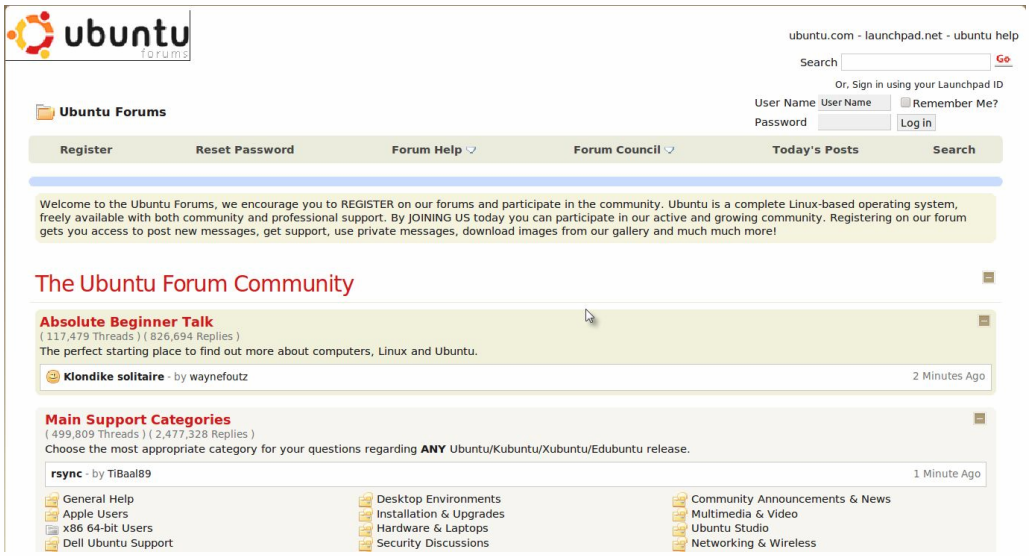

Fig.16. 4: Ubuntu Forums: Online bulletin board for Ubuntu users. (http://ubuntuforums.org/)

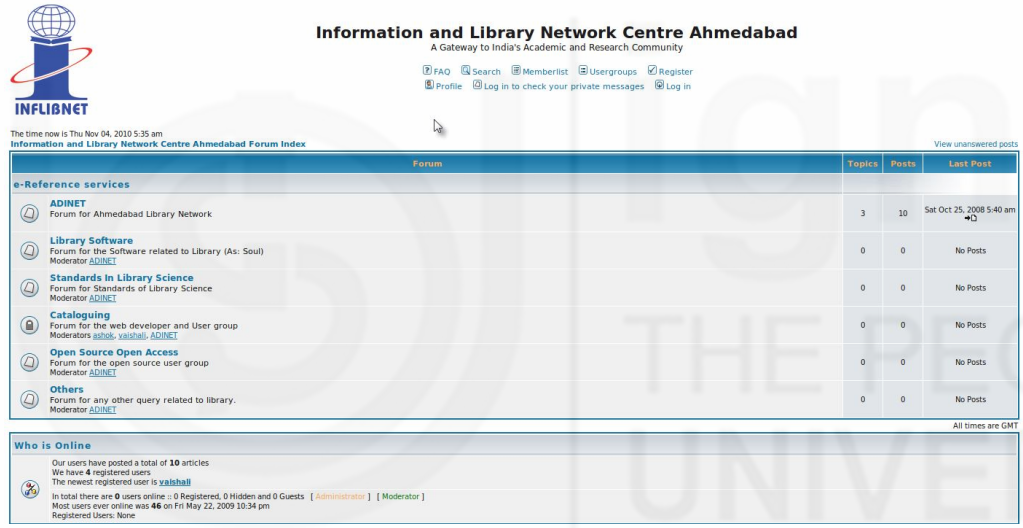

Fig. 16.5: INFLIBNET Bulletin Board (http://www.inflibnet.ac.in/forum/)

16.2.3 Mailing Lists and Discussion Lists

Mailing lists, discussion lists or listservs are services facilitate sending e-mails to a group of individuals with ease. These different names refer to the same process whereby one can send e-mail to a large group of people rather like using CC functionality provided by various mail applications. They are usually fully or partially automated through the use of software such as GNU's Mailman, Listserv, Mailbase etc. This service is hosted on hosting server that provides a reflector address on the same server capable of receiving email. The hosting service also maintains a list of all the different Mailing lists and the people who subscribe to the lists. The software processes incoming messages sent to the reflector address and depending on the content of the messages they are acted upon internally (in case of messages containing commands directed at the software itself) or are distributed to all e-mail addresses subscribed to the mailing list (See fig. 16.6). Joining a mailing list is called "subscribing" and leaving a list is called "unsubscribing".

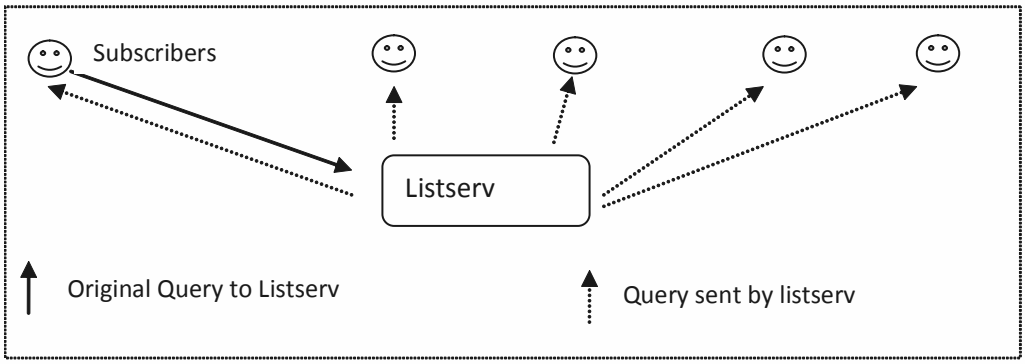

Fig. 16.6: Diagram Showing the Operation of a Mailing List/Listserv

There are thousands of mailing lists available on Internet; these are devoted to a varied range of topics and individuals. The popular open source developments and W3C standards first open a public mailing list to have ideas and to discuss on the standard in hand and after the discussion the changed are made to the standard. The best use of mailing list is for discussion, some libraries use mailing list to alert the users about new arrivals or sending the table of contents of new journal issues. In general mailing lists provide a forum to:

- Get advices on buying new systems.
- Request any factual information from the group.
- Support the group members in solving problems.
- Announce upcoming meetings and conferences.
- Inform about new vacancies.
- Discuss some prevailing issues faced by the profession.
- Information about new websites, productions and publications.

Some mailing lists have the functionality to send mails in two modes viz. digest and mail-as-arrives. The digest mode is good for heavily active lists. Digest is a consolidated mail containing string of mails sent by mailing list. This helps users' who wants to look all the mails in one stretch and does not want to be disturbed regular receipt of mail. Users' can fix the frequency of digest mail once in day or two or the way they want. Otherwise users' receive mails regularly as the mail is shot by mailing list. Some examples are:

1) **LIS Forum:** largest mailing list comprising mostly library and information science professional from India. It is operated and maintained by NCSI, Indian Institute of Science, Bangalore. (http://ncsi.iisc.ernet.in/mailman/listinfo/lis-forum)

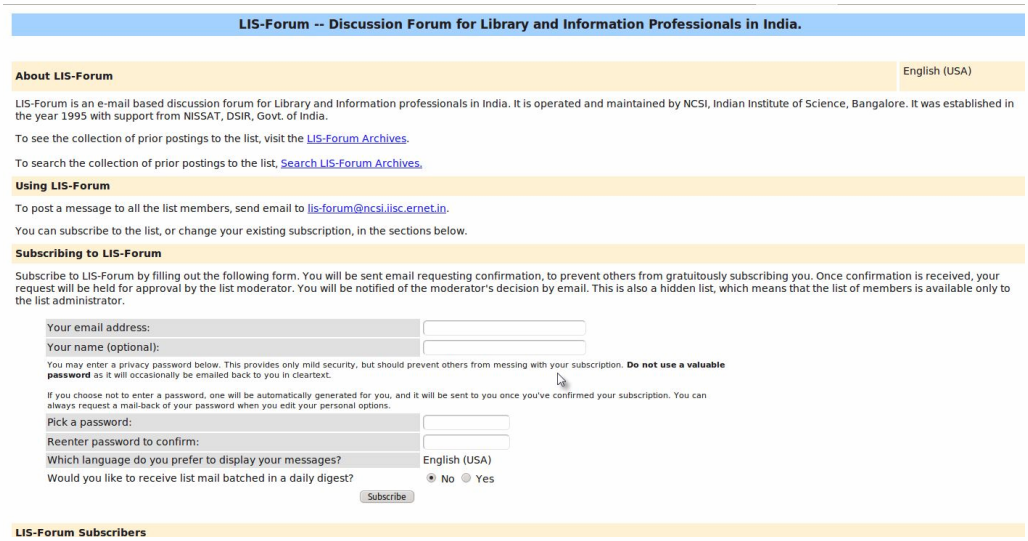

Fig. 16.7: Mailing List of NCSI, Indian Institute of Science, Bangalore. (http://ncsi.iisc.ernet.in/mailman/listinfo/lis-forum)

2) **DLRG:** Digital Library Research Group is a mailing list for helping library professional in solving software and hardware related problems. This is operated and maintained by Documentation Research and Training Centre (DRTC), Bangalore. It mainly covers topics related to digital library, library management systems and other data import and backup.

Fig. 16.8: Digital Library Research Group of DRTC, Bangalore

16.2.4 Resource Sharing

Resource sharing is a partnership where several libraries share one or more of their functions, for example, acquisitions, processing, storage and delivery of services. Each member has something useful to share, is willing to share and a plan exists to accomplish this. Major goal of resource sharing is to augment the local holdings by providing access to collections of other libraries.

One of the major approaches of resource sharing is sharing of library catalogues. For this purpose, libraries use Z39.50 protocol. In this environment, different libraries can be searched with one single interface. This interface can also be used to download the library catalogue entry so that duplication of work can be avoided among the libraries.

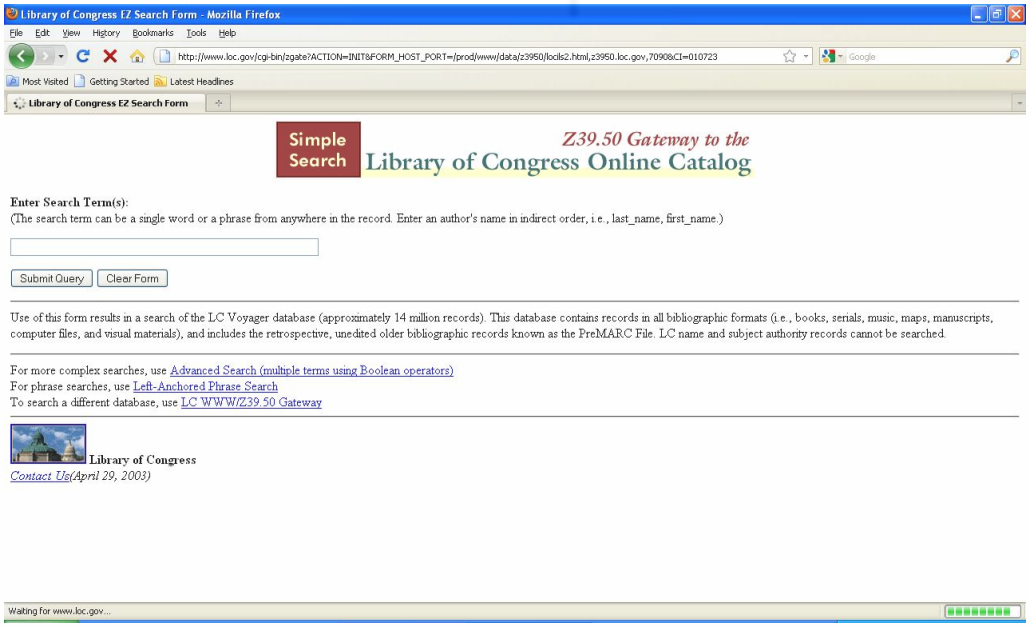

Fig. 16.9: Library of Congress Z39.50 Search

16.2.5 Online Document Repositories

Document repository systems are digital libraries containing an organisation's documents. They are commonplace for all the important documents. These systems allow controlling how documents are created, accessed, stored, and even disposed of. These may look similar to a file server but they have many differences such as. It allows creating templates for a particular class of documents for example, one can create a template for user requests wherein the user or staff may feed data wherever required and in the case of standard information the data will be prefilled. This ensures the consistency in data entry and ease in data entry.

Wherever there is sensitivity of information contained in the documents, the repository will need to be secured. For this the technology of encryption and authentication mechanisms is to be used. The online document repositories are popular in corporate world. In big organisations the employees have their username and password to access the system.

Some of the institutions and university libraries are also using document repositories to store and provide access to their research output.

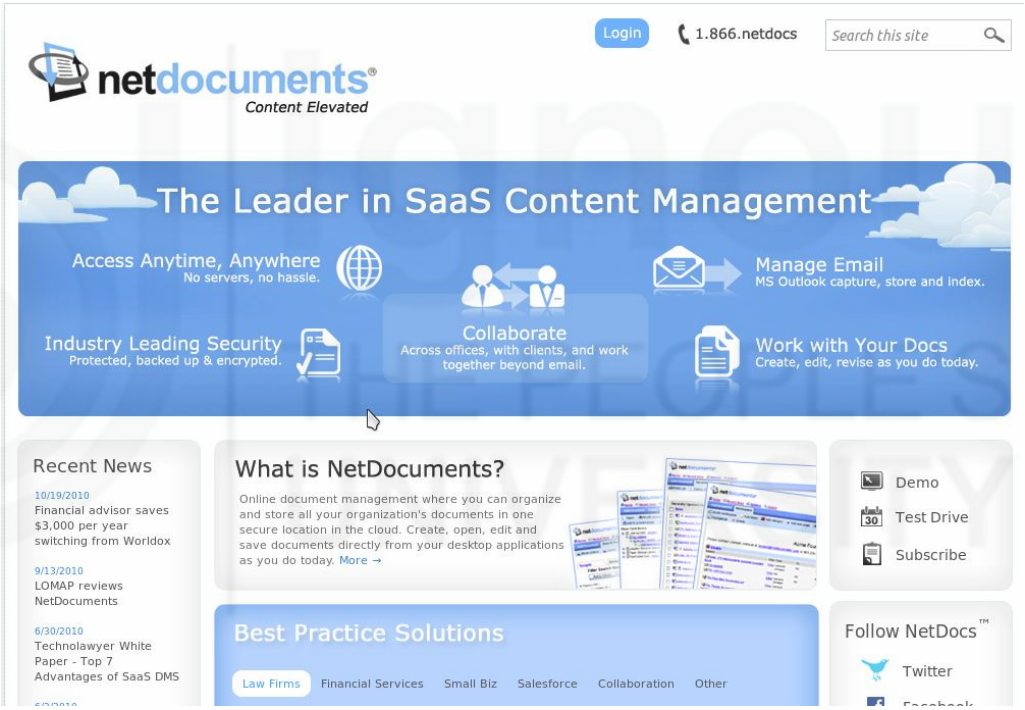

Fig. 16.10: Netdocuments (http://www.netdocuments.com/)

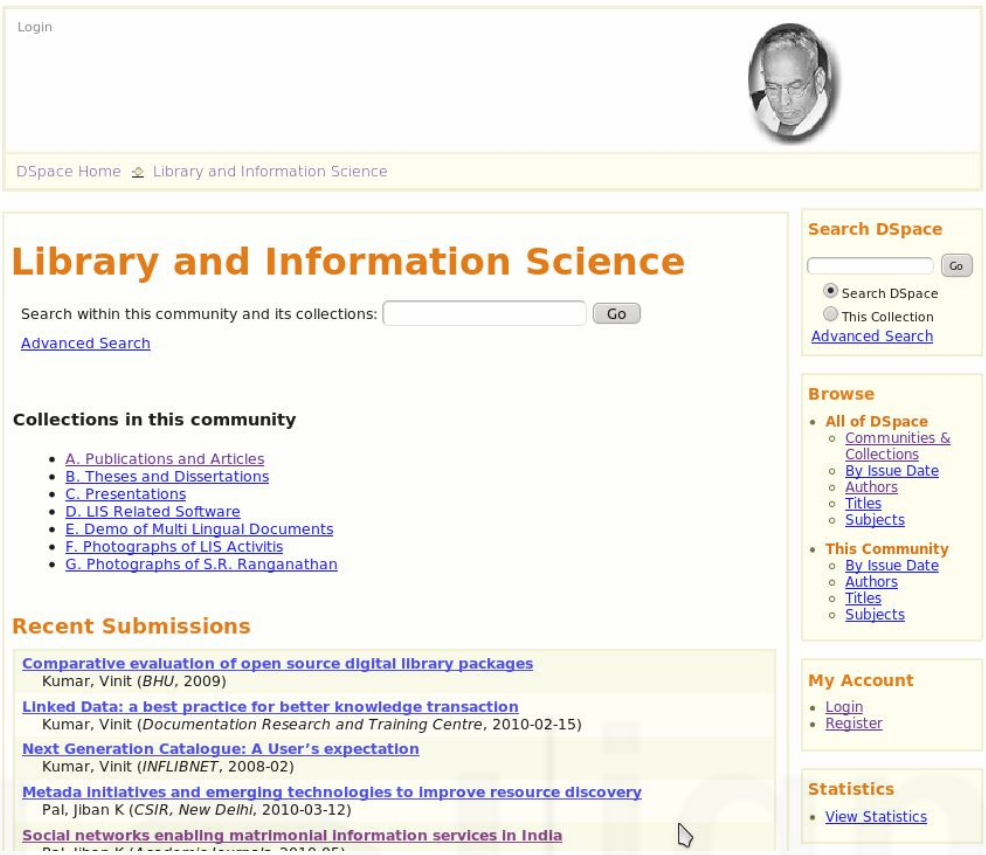

Fig. 16.11:Librarians Digital Library (LDL)(https://drtc.isibang.ac.in)

16.2.6 Web Portals

Web portals are a gateway to the information services on the Web or other sites on the Web. These act as a convenient location of sites of related interest, as it is seen in subject specific portals like infolibrarian.com or in general portals like yahoo.com. These portals provide other services such as e-mail, news, stock prices, information, databases and entertainment.

A truly effective portal must include:

- A single point of access (Single sign on)
- Unified search across all information sources
- Personalisation
- Applications integration
- Collaboration
- System security
- Openness
- Links to help files

Fig. 16.12: Yahoo Web Portal (www.yahoo.co.in)

There are subject specific web portals. These portals include documents and resources within the scope of subject. An example, in library science is infolibrarian.com.

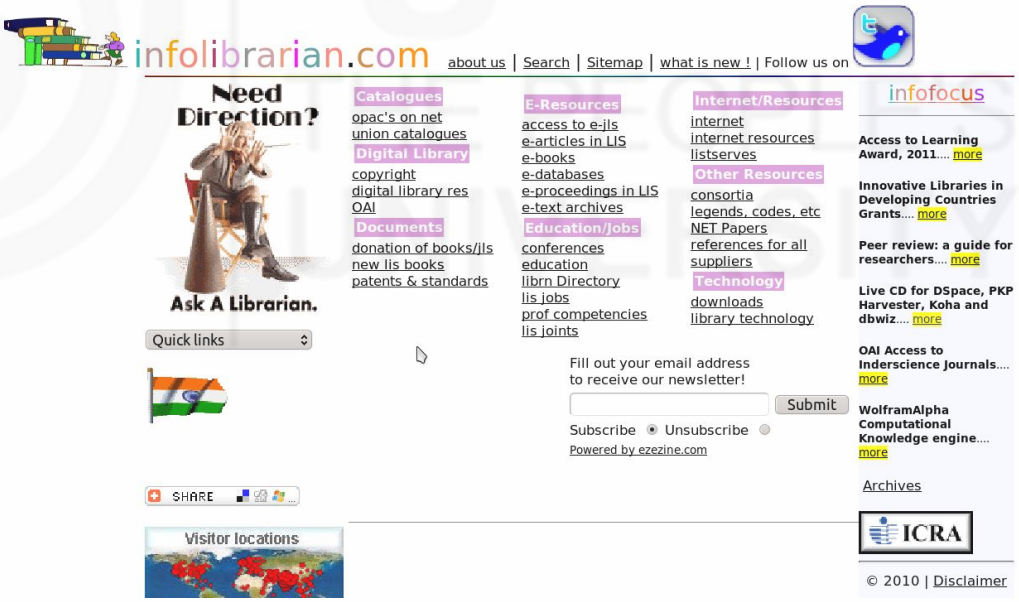

Fig. 16.13: Infolibrarian.com (http://www.infolibrarian.com/)

16.2.7 E-mail

E-mails are very common and popular method of exchanging information. E-mails are commonly and regularly used for both formal and informal communication. It is also used as a popular means of keeping up-to-date and solving small queries that arise everyday in the workplace.

International Encyclopedia of Information and Library Science defines e-mail as "a method of sending messages, data files, etc. by electronic means from one computer

with network access to another". The receiving server or machine is usually equipped with a storage area, or mailbox, in which the messages are deposited. The access to mailbox is restricted by password and only the authenticated users having the password can read the mailbox. Users can read their incoming messages on-screen when they choose and, if they wish, print them out or download them on to a disk. In the beginning, the e-mails were delivered, as user to user which required both the users to be online. The present day e-mail servers provide users with the capabilities to store and forward the messages, which mean that the need to be online is not required. Further the users can login to the server and can check the messages at any time. The messages are exchanged between the users through a protocol known as Simple Mail Transfer Protocol (SMTP) with software programmes called Mail Transfer Agents (MTA) such as Exim4, Postfix and Sendmail. The MTAs provide user to retrieve the messages through standard protocols viz. Post Offices Protocol (POP) or Internet Message Access Protocol (IMAP) some corporate organisation have some proprietary standards such as Lotus Notes and MS Exchange server. The messages are written using clients which are desktop based applications such as Thunderbird, MS Outlook, Apple Mail, etc. or web based such as Webmail etc.

Some of the existing services and clients on Internet:

Fig. 16.14: Gmail - Google's Mail Service: (http://mail.google.com)

Fig. 16.15: Organisational Mail Service

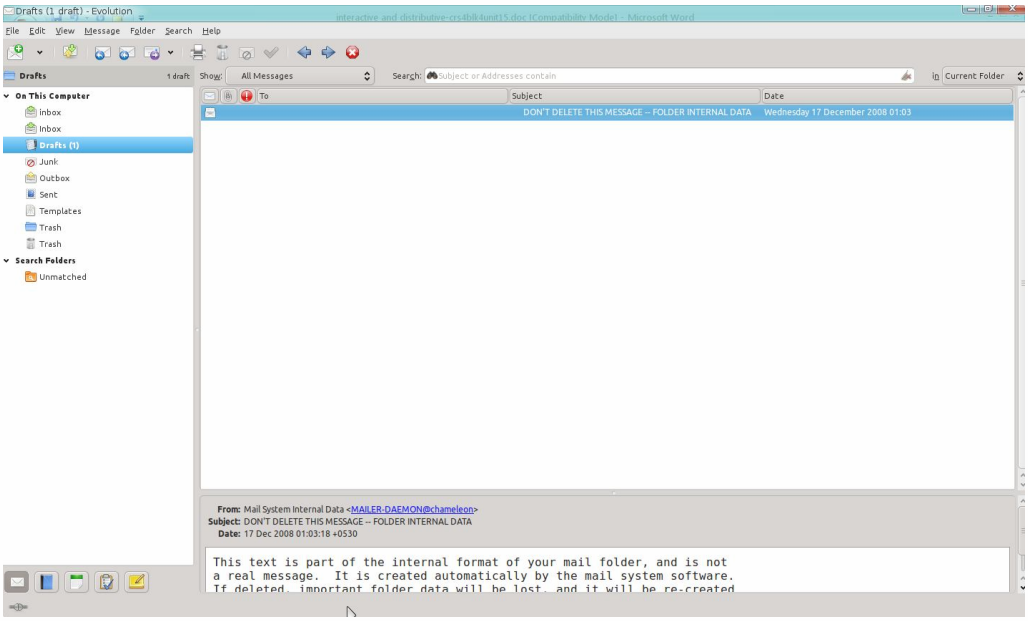

Fig.16.16: Evolution – Default Mail Client in Ubuntu Linux Distribution

The email services have some advantages and disadvantages too. The advantages are:

- The speed for delivery of messages is quick and almost instantaneous.
- They are reliable and secure efforts are going on to make them more reliable and secure.
- They are environment friendly, as there is no use of papers. Similarly the messages are stored permanently so it is easy to find the old messages.
- The use of graphics such as a picture etc. adds value to the service. Because of its one-to-many nature.
- It can be used as publicity and advertising tools as well.

Some disadvantages are also associated with email service such as:

- Because of its electronic nature and the facility to send attachments there are chances for viruses to get distributed through it.
- There is a big problem of spamming, which is nothing but sending unsolicited emails or advertisements. Checking and deleting these unsolicited emails can unnecessarily consume a lot of time of users.
- As the system is password based and contains lot of private data it gets more prone to security threats.

16.2.8 Online Storage and Searching

The electronic systems are very prone to mishandling and chances of getting corrupt due to myriad of reasons such as virus attack, natural disaster, theft, lost or broken, etc. Some online services on Internet provide the capability to store the files for easy retrieval and backup. On top of it they provide features like:

- Access files from any computer and anywhere, which reduces the need to carry them physically or sending them to own e-mail address.
- Some services allow viewing and listening media like pictures and audio files.
- Provide easy way to share the uploaded files among friends, family or the world.
- Remote access to the files, via a desktop client or website.

These services are also encrypted and properly authenticating looking at the sensitivity

of the data. These services are mainly priced, but for small storage space they are free. The clients are seamlessly integrated to the operating system of the user. Some services (like, dropbox.com) dedicate a folder in the users' computer and allow dragging and dropping any file in the folder automatically synchronous to the online server. Similarly, some services provide programs to schedule the time of online syncing. Some of the online file storage services are listed in the following table.

Table 16.2: Online Storage and Searching Services

Service Name	URL	Priced/free
Dropbox	www.dropbox.com	Free till 8GB
SugarSync	https://www.sugarsync.com/	30 Day trial
OpenDrive	http://www.opendrive.com/	Free 5GB
Mozy	http://mozy.com/	Free till 2GB
Box.net	http://www.box.net/home	Free till 5GB

Apart from these exclusive storage services we have services like Google Drive (https://drive.google.com/ý) and Microsoft SkyDrive (https://skydrive.live.com/ý) wherein one can use existing e-mail accounts to use the storage facility. These services are also available on mobile devices like smart phones and tablets.

16.2.9 E-publishing

The information produced is disseminated by different model of publishing. New technologies have transformed the process of publishing and distribution of information. Electronic publishing is the process for production of typeset quality documents containing text, graphics, pictures, tables, equations etc. in digitised form. It uses new technology allowing publishers to deliver documents and other contents quickly and efficiently as well.

There are two important modes of e-publishing i.e., online and offline. Online publishing is in the form of online journals, websites, online database, e-books and so on. Offline publishing is done over some storage media like, Compact Disc, DVD and so on. Now-a-days e-books are being published on DVD and CDROM.

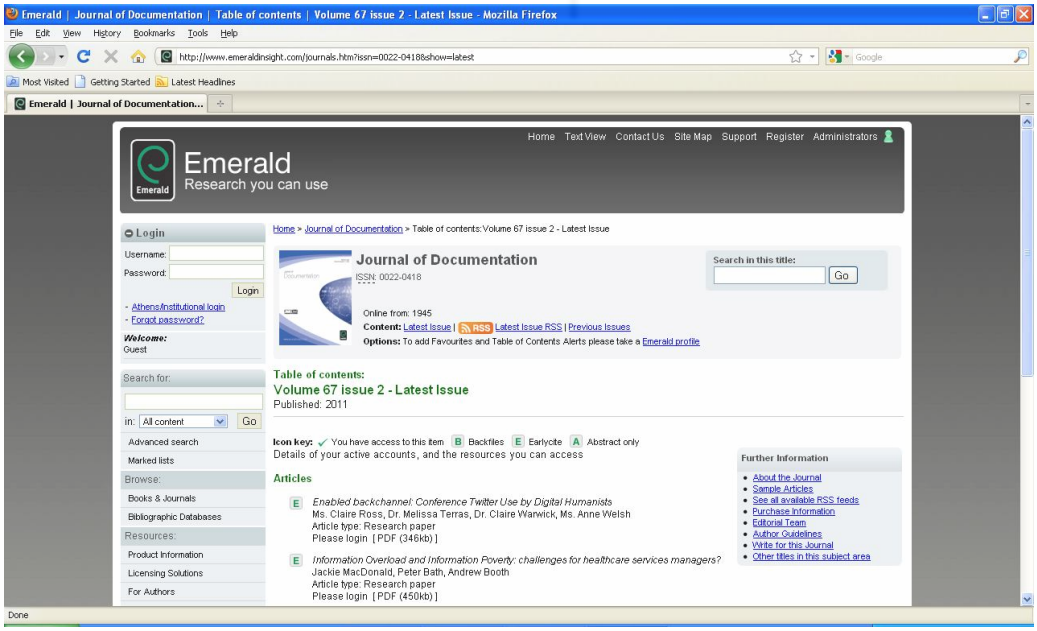

Fig. 16.17: Emerald Online Publishing

16.2.10 Webcasting

Webcasting is a service based on push technology, which means updating the systems by which users of the Internet, or Intranets, can receive news and other information through periodic and unobtrusive transmissions. The webcasting is an easy way to distribute the relevant content to the users. A webcast is a media file distributed over the Internet using streaming media technology to distribute a single content source to many simultaneous listeners/viewers. Webcasting is the Web-enabled broadcasting and integration of dynamic rich media. In simpler terms webcasting is "broadcasting" over the Internet. This involves one-to-many communication type. Webcasting informs users about material relevant or related to their interests before any specific request is made. This is done on the basis of pre-selected topics stored in a user profile. The user profile is further updated on the basis of the received feedback. The websites selected are referred to as channels; the updates received from the channels are supplied in its raw form or sometimes it is sent after processing.

With the use of very cheap and accessible technology it has become very easy to webcast anything such as:

- Meetings
- Special events: Birthday celebrations, festivals etc.
- Collaborative engineering
- Conferences with key partners
- Discussions with supply chain partners
- Sales presentations and demonstrations
- Voting, panel discussions
- Interviews
- Keynote addresses
- Training

There are many instances of webcasting available now-a-days with topics covering technology, computers, news, cooking, tutorials etc. The news webcasting is very popular these days. Webcasting is sometimes referred as "buffering media" too. Some examples of webcasting are:

1) **YouTube**

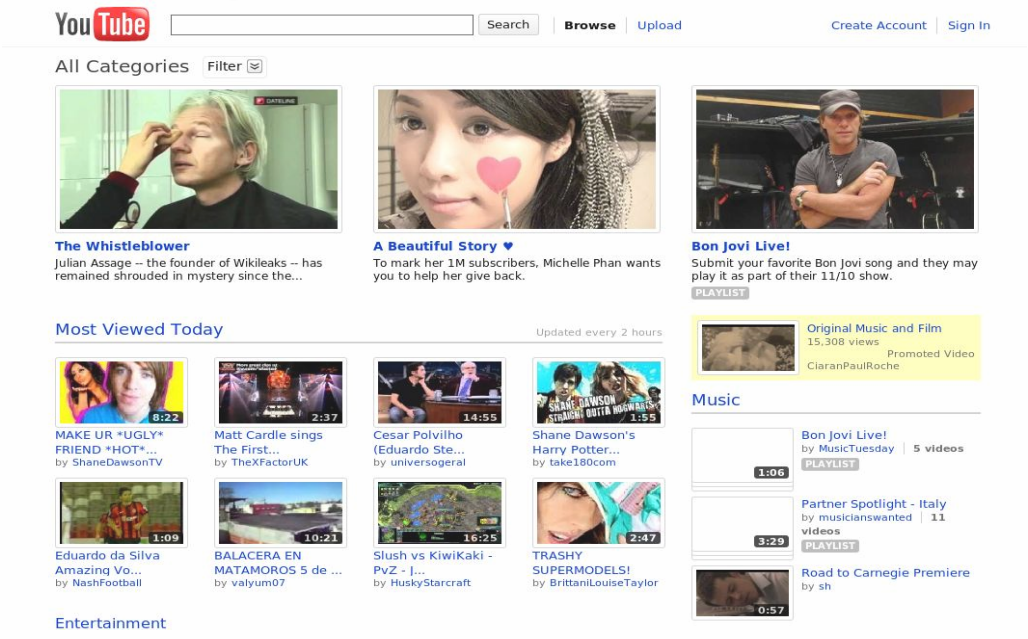

Fig. 16.18: Webcasting through YouTube(http://www.youtube.com/)

2) **BBC World News**

Fig. 16.19: News Casting(http://www.bbc.co.uk/iplayer/console/bbc_world_service)

3) **IGNOU Gyan Darshan**

Fig. 16.20: IGNOU Gyan Darshan Channel (http://www.ignouonline.ac.in/Broadcast/)

Self-Check Exercise

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

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

1) What is Web Directory?

2) Discuss advantages of Email communication.

.....

.....

.....

.....

.....

.....

.....

.....

16.3 INTERACTIVE DISTRIBUTED SERVICES

With the implementation of Web 2.0 technology the services are not only distributed but they are interactive also. They are more users' centric or customer oriented. The implementations of interactive services are manifold. Some of these are discussed in the subsequent sections.

16.3.1 Interactive Learning

Networked technologies such as Internet and World Wide Web are dramatically changing education and training as they enable people to access information and communicate with others across terrestrial boundaries, cultures and on a global scale. Interactive learning as defined by Chartered Institute of Personnel and Development emphasises the importance of connectivity and interactivity states "Learning that is delivered, enabled or mediated by electronic technology, for the explicit purpose of training in organisations. It does not include stand-alone technology-based training such as the use of CD-ROMs in isolation". Use of interactive learning involves text, graphics, audio, video and animation. Apart from this, the programmes are enhanced by providing additional support, e.g. using synchronous and asynchronous communication applications such as e-mail, discussion groups, chat rooms and video conferencing. The learning usually takes place through the use of web-based training programmes, where the learner typically follows a pre-specified learning process that includes opportunities for practice and assessment and feedback activities. It also takes place through blended approaches that involve learners experiencing a mixture of face-to-face and online learning experience.

There are online learning environments both on commercial and open source platforms. Moodle is one example of hosting online e-learning programmes.

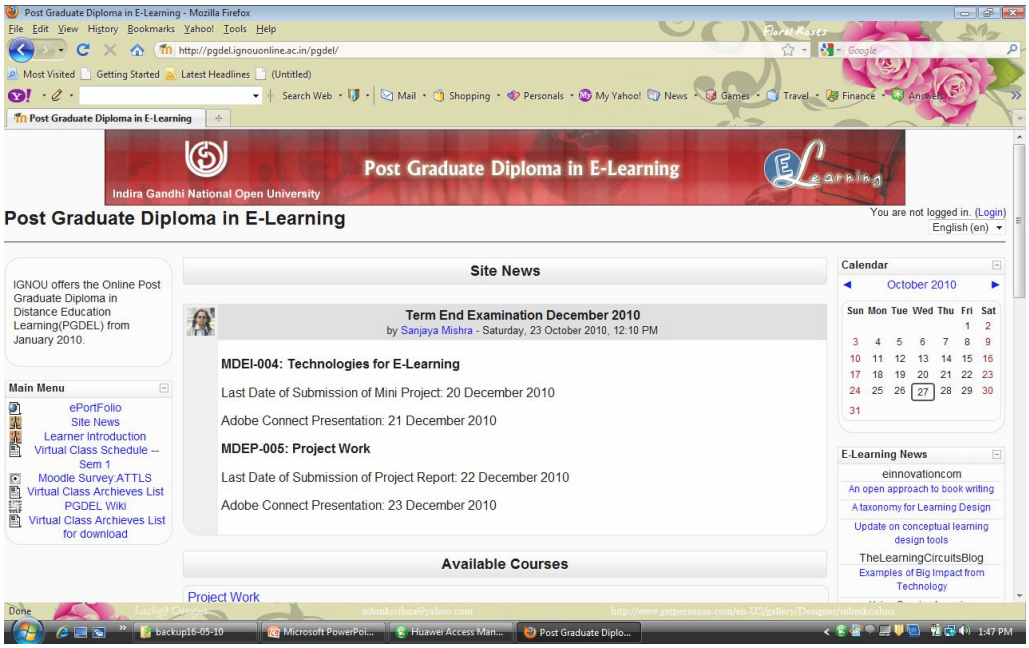

Fig. 16.21: Elearning Programme Using Moodle LMS

16.3.2 Interactive Business and Trading

One of the major implementation of existing web technology is seen over online trading and business. With the evolving standards and inter-operability among systems business houses are extending their sale counter over Internet. A user can order the products and make the payments online. An important interactive service over such online shops is use of cart facility where customer can collect the products s/he wants to buy and at the end make the payment. These carts are interactive tools where one can add or remove the product as one does in real stores.

The interactive environment of business has been extended to booking tickets online with selecting the best available option to travel. Online trading of live stocks in share market is another area which has been revolutionised. The live trading of stocks can be done. The trading window displays live prices of stocks every second. Same kind of trading services are also offered by many financial banks online where a client can buy and sell stocks and immediately money is deducted and deposited in the account as the case may be.

Fig.16.22: Online Stock Trading Screen

(Source: http://www.aaasec.com.my/stocksen/quicktour3.asp)

16.3.3 Remote Computing and File Transfer

Remote computing is a regular phenomenon in networked environment. It is used to share expensive hardware and software among a group of people. For example, if one has to perform certain set of analysis on a certain set of data. S/he can transfer the file using File Transfer Protocol (FTP) from present machine to remote machine where such analysis facility is available. Thereafter, whole processing can be done on remote machine using the resources from the remote machine. In such cases, the local machine acts as a dumb terminal.

With the increased use of Internet and implementation of new technologies, the form of remote computing has changed to Cloud Computing. In cloud computing the term cloud represents the use of Internet. The resources in cloud computing are distributed and often the user of computing facility is also not aware fully of location or the machine which is offering service. The cloud computing may use more than one resource distributed at different locations far apart.

Cloud computing is a way to reduce the cost on expensive computing exercises. It offers a sustainable environment for users to use and share their costly devices with others. One of the major implementation of cloud computing is Beowulf cluster. These are set of inexpensive computers, which are collectively used to process data as high-performance parallel computing device. These are shared machines working parallel over network.

16.3.4 Interactive Communication

Interactive communication is an important factor for the innovation and collaboration in present ideas hunting world. There are some prominent and powerful tools available to help to do so such as Microsoft Exchange (http://www.microsoft.com/exchange) server. It is a server based system and provides users with collaboration tools similar to Microsoft Outlook e-mail, calendaring application. Exchange provides shared calendars, shared address books, e-mail and other collaboration tools. The MS Exchange user typically accesses an Exchange account via a web interface or the Exchange desktop application.

There are tools, which are used for interactive real-time chat either through text or for conferencing. The participants can interact in real time to share their views and provide feedback or answer immediately. We will now discuss some of the live chatting and web conferencing tools.

Some of the tools for live chat are as follows:

1) Microsoft NetMeeting (http://www.microsoft.com/windows/netmeeting): It provides real-time chat, file transfer, interactive whiteboard and file sharing. NetMeeting also provides face-to-face video conferencing using webcam or other source and real-time audio communication.

2) Windows Messenger: Messenger allows for sending instant messages to other Windows users, audio, video, communication, file exchange, text based chat, and integration with 'remote assistance' to allow remote login to the local PC by a technical support service. Messenger is provided with most versions of Microsoft Windows.

3) ICQ (http://web.icq.com) This application provides similar functionality to Windows Messenger and may also be used via a web browser interface at the ICQ website. ICQ is a worldwide chat system, and should be used with care in any serious context.

Tele/video conferencing is the need of the day particularly on the business and trade. Facility of conducting online tele/video conferencing reduces the burden (financial and physical stress) for the business houses. This has come up as a major solution where long distance travel is required. It has become an interactive medium to communicate real time face to face. People traveling at far off place can keep in touch with their family and friends. Normally, computer-to-computer communication is free however service for calling a telephone is paid which is very nominal. This has revolutionised the global communication scenario.

The conferencing can be done one to one basis or among a group of nodes simultaneously. Some popular services are listed in Table 16.3.

Table 16.3: Conferencing Tools

Application Name	Features	URL
Pidgin	Same as Window Messenger	http://www.pidgin.im/
Skype	Similar features as NetMeeting, but more popular	http://www.skype.com/intl/en-us/home
Google+ Hangouts	Online chatting-audio, video and text chatting	http://www.google.com/hangouts/
Yahoo Messenger	Online chatting-audio, video and text chatting	http://messenger.yahoo.com

16.3.5 Interactive Search Agent and Document Delivery

Though it is in a very nascent stage but there are services, which are of interactive nature. Interactive search agents are normally domain specific. One of the implementation of domain specific search agents is seen in travel booking where a single search result yields many options. The user can choose which ever suits them and further refine the search and book for their travel.

Similarly, Google and Yahoo have launched a search service to make search more interactive and user oriented. The search interface helps users to find most suitable search term in order to form their search query. While typing the search term in the box all the related search terms are also displayed in anticipation to help user to find the search query.

The search engines provide facility to preview the searched pages within the search result page so that users can gaze the importance of web page towards their search and save their valuable time. Google has launched personalised searching alerts for any new addition by email. The service is known as Google Alert. Search engines do provide option to search within the blog, webpages, or statistics and so on.

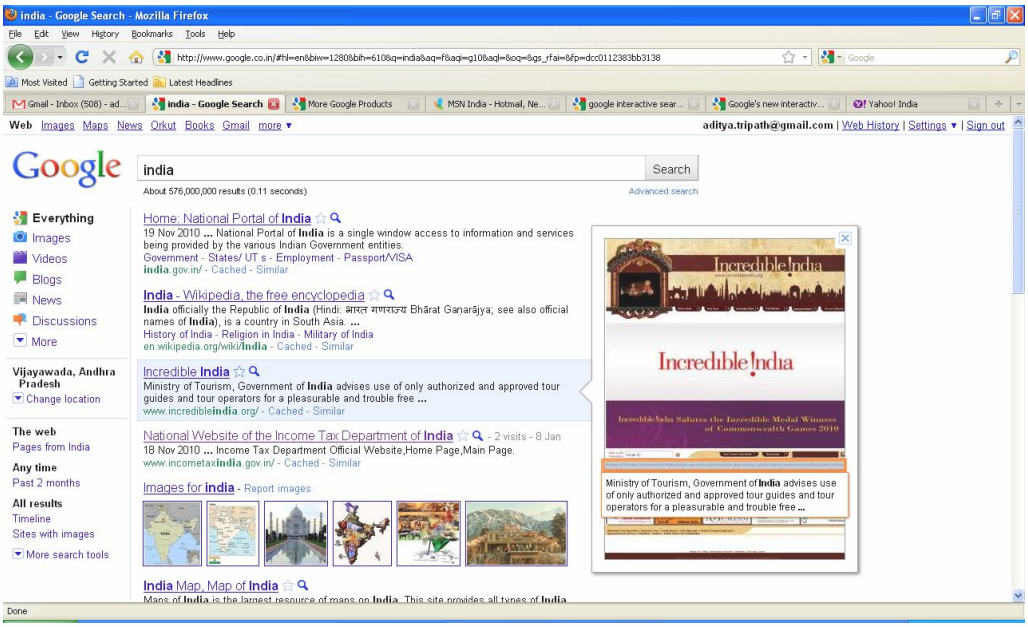

Fig. 16.23: Interactive Web Searching by Google

16.3.6 Interactive Bookmarking

People keep working online and come across several resources. These resources are book marked for any future use. The online bookmarking services provide method to bookmark resources online and use wherever required. One of the examples of such bookmarking is Google Bookmarks. The bookmarks can be categorised online and labeled.

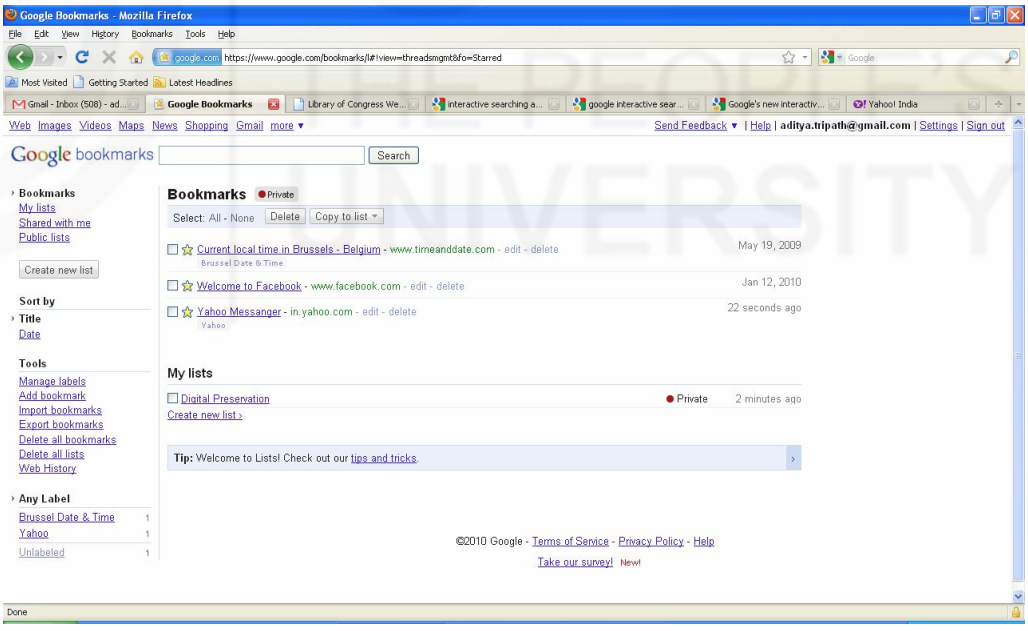

Fig. 16.24: Google Bookmark

16.3.7 Interactive Translation Service

There are several websites, which provide online translation service. However, the machine translation is not a matured technology but there are several implementations to it particularly in the web environment.

The tools like, Yahoo Babel Fish translates online given text and webpage. Google presents language tools for translation. It supports Indic language translation.

Fig. 16.25: Google Translation

Self-Check Exercise

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

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

3) What do you understand by Interactive Web service?

.....

.....

.....

.....

16.4 SECURITY AND PRIVACY ISSUES

Security is always an important issue in the design of any interactive and distributive service. Internet security means to prevent any unauthorised access to networked resources and services. Typical methods for ensuring security include:

1) **User authentication:** Integrating the service with a user records database (such as Active Directory) and an authentication system, such as LDAP (Lightweight Directory Access Protocol) should allow for use of the normal network username and password to access the services, restricting access to individuals without and institutional network account.

2) **IP:** The IP is a unique numerical network address on every computer. Access to the implemented services may also be limited using IP restriction, i.e., limiting the access to a range of known computers, probably located within an organisation or distributed over location.. For distributed locations the IPs of the machines should be known to the system running the IP restrictions.

3) **User roles:** The provision of having access control lists (ACLs) to define the roles of the registered users provides good level of security and reduces the chances of abuse. Some roles such 'system administrator' to manage the hardware and software at the highest level can be created. Similarly, service specific roles like 'moderator' etc. can be created for moderation of chats and discussions. Use of roles is an important security consideration, because higher level roles provide access to complex features, which if used improperly could damage the system or result in data loss (e.g. removal of user accounts).

Similarly there are privacy issues in communicating over a network, since all

communications can be intercepted and recorded without participants being aware. The distributive services and social network present nowadays on web are more concerned with the privacy of the users. There are reports that the user information available in some of the services were mishandled and reached in the hands of marketing companies. These companies use the users data to propagate their products and to do business analysis etc. This further is breach of privacy of the user, as the offers offered by these companies will bug the users personal life. The Encyclopedia for Library and information Science defines privacy as "The quality of protection for aspects of the life, and information about the life, of an individual or a group, from the intervention or knowledge of others". The handling of privacy is under the eyes of media also, so utmost care should be taken. Such as:

1) **Privacy policy:** Every service asking users about their personal details should draft a privacy policy. Respect for the privacy of certain aspects of information is a counterbalance to the principle of freedom of information.

2) **Encryption:** Since all the communication in interactive and distributive technologies is transferred over a network these can be intercepted and recorded without the knowledge of the persons involved. Encryption technologies especially the public key encryption system should be used to keep the communications coded and readable only by the intended participants.

3) **Moderation policy:** When a system provides the interaction platform it may happen that some users may abuse the platform, especially in the case of discussion forums and chats. This is very important issue as the improper handling will not only refrain the fair users it will also make the environment nasty as well. To limit the abuse some moderation policy should be framed and implemented with some moderators. The steps to disallow the anonymous postings, banning of such kind of users etc. can be framed in the moderation policy. Similarly for chat, the sessions may usually be recorded providing a text record of all messages.

16.5 SUMMARY

In the present Unit we studied that Web has become a medium to deliver different types of services. The mode of many of the old services has changed to be more interactive and user oriented. Most of these services available are free and users can use them in the format they desire and as per their requirement.

We also learnt about different types of distributed services like bulletin board, forums, web portals, web casting and so on in this Unit.

The easy access to plethora of interactive web services has brought up several issues like data privacy and encryption, which has been discussed as well.

16.6 ANSWERS TO SELF-CHECK EXERCISES

1) Web directory is a topical list of Internet resources arranged in a hierarchical way. It is organised Web site listings created by human reviewers and act as a search tool created by human. It is quite different from search engine and does not display lists of web pages based on keywords; instead, it lists web sites by category and subcategory. The categorisation is usually based on the whole web site rather than one page or a set of keywords, and sites are often limited to inclusion in only a few categories.

2) Following are the advantages of E-mail

- The speed for delivery of messages is quick and almost instantaneous.
- They are reliable and secure and now efforts are on to make them more reliable and secure.
- They are environment friendly, as there is no use of papers. Similarly the messages are stored permanently so it is easy to find the old messages.
- The use of graphics such as a picture etc. adds value to the service. Because of its one-to-many nature.
- It can be used as publicity and advertising tools as well.

3) An interactive web service is online service given by a service provider or agent. In an interactive service user can provide input or feedback to the service provider. He can also modulate the mode of delivery of service or interface of service environment if he is permitted to do so.

16.7 KEYWORDS

Bookmark : A *bookmark* is a locally stored Uniform Resource Identifier (URI).

Database : A *database* is an application that stores and organises data for fast retrieval of information.

Digest : A style or format of distribution of electronic mailing lists in which multiple messages are placed together and distributed as a single unit.

E-publishing : Publishing of e-books and electronic articles over web to establish digital libraries or build organisational repositories.

Encryption : Coding of data for secured transmission.

Forum : An online group where visitors may read and post topics of common interest.

Mailing Lists : A mailing list is an electronic discussion forum that anyone can subscribe to. When someone sends an e-mail message to the mailing list, a copy of that message is broadcast to everyone who is subscribed to that mailing list.

MIME : **It stands for** ***M****ultipurpose* ***I****nternet* ***M****ail* ***E****xtensions. It is* a standard for formatting non-ASCII messages so that they can be communicated over the Internet.

Network : A group of two or more computer systems linked together.

Online document repositories : Document repository systems are digital libraries containing an organisation's documents.

Personalisation	:	User-specific tailoring of information.
Portal	:	These web sites serve as a single platform to access content and applications on a topic. It is also known as a "gateway," a "research guide," a "virtual library" or a "Web directory."
Query	:	A request for information from a database.
Remote Access	:	In web parlance, it is defined as access to a computer or a network from a remote distance.
Repository	:	A place where multiple databases or files are located for distribution over a network.
SMTP	:	Simple Mail Transfer Protocol is a set of standards to define the process of exchange e-mails between two servers.
Thread	:	A posting (message) on a discussion group or a mailing list and all of the responses to it builds a *thread*. To "follow a thread" is to read a series of messages sharing a common subject.
Thunderbird	:	Free, open source, cross-platform e-mail and news client based on Mozilla code.
Web Directory	:	A search tool created by editors or trained researchers who categorise or classify Web sites by subject.
Webcasting	:	A web-enabled broadcasting of information to the users.

16.8 REFERENCES AND FURTHER READING

1. Allan, Barbara. *E-learning and teaching in library and information services*. Facet publishing: London, 2002. Print

2. Burchard, Paul. *Towards Standards for an Interactive Web*. Available at <http://www.geom.uiuc.edu/hypernews/get/interactive/index.html>

3. Casey, Michael E. and Savastinuk, Laura C. Library 2. 0: A Guide to Participatory Library Service. *Information Today*: New Jersey, (2007), 174 p.

4. Catherall, Paul. *Delivering E-learning for information services in Higher Education*. Chandos Publishing: Oxford, 2005. Print

5. Chad, K and Miller, P. *Do Libraries Matter: The rise of Library 2.0*. 2005 <http://www.talis.com/applications/downloads/white_papers/DoLibrariesMatter.pdf>

6. Chapanis, Alphonse. Interactive Human Communication. *Scientific American*, 232 (1975) p.36-47. < http://web.media.mit.edu/~geek/class/ChapanisSciAm.pdf >

7. Cohen, Laura B. (ed.). *Library 2.0 initiatives in academic libraries.* Association of College & Research Libraries. (2007) 167 p.

8. Crichlow, Joel M. *Distributed Systems: Computing over Networks*. Haryana: PHI Learning, 2009. Print

9. da Silva, Alberto Manuel Rodrigues. Tutorial on web services. <http://cita2003.fing.edu.uy/articulosvf/WS1.pdf>

10. Edwards, Jim. Web Casting - The Future of Online Education. <http://www.teachnology.com/tutorials/webcast/ >

11. Holmberg, Kim, et.al. What is Library 2.0?. *Journal of Documentation*, 65.4 (2009) pp.668 - 681

12. Interactive Learning Opportunities On The Internet <http://www.techteachconcepts.com/interactive_learning.htm>

13. Internet Relay Chat (IRC), < *http://www.livinginternet.com/r/ri_early.htm*>

14. John-Okeke, Rita. Web-based legal information services and academic law libraries in Nigeria. *Library Hi Tech News*, 25. 9 (2008) pp.12 - 16

15. Khosrow-Pour, Mehdi (ed.). *Advanced topics in electronic commerce*. Idea Group, 2005. Print

16. Kidd, Terry T. and Chen, Irene. *Wired for learning: an educator's guide to web 2.0.* Information Age Publishing, 2009. Print

17. Kovacs, Diane K. "Electronic publishing in libraries: introduction", *Library Hi Tech*, 17.1(1999), pp.8 - 9

18. Librarians and the Internet, Social Media, and Web 2.0 <http://www.library20.org/ >

19. Mailing Lists, < http://www.livinginternet.com/l/lp.htm>

20. Maness , Jack M. "Library 2.0 Theory: Web 2.0 and Its Implications for Libraries" *Webology*. 3.2 (2006) <http://www.webology.ir/2006/v3n2/a25.html>

21. Open Source Document Repository, <http://www.vclcomponents.com/s/0__/open_source_document_repository/>

22. Salber, D., Privacy Issues in Computer-Mediated Communication. *Science-and-Engineering-Ethics*, (1996)p. 238-240.

23. Shou, Joe. A history of Web portals and their development in libraries. *Information Technology and Libraries*. (2003) <http://www.accessmylibrary.com/article-1G1-109847209/history-web-portals-and.html>

24. Steinbrenner, Karin "PORTALS", Evolving Technology Committee. (2001) <http://www.humane.eu/fileadmin/wsan_docs/ws2004/ws2004_Fri_Veling_1_Educause.pdf>