
UNIT 1 CONCEPT AND SCOPE

Structure

- 1.0 Objectives
- 1.1 Introduction
- 1.2 Types of Documents
 - 1.2.1 Books
 - 1.2.2 Journals/Articles
 - 1.2.3 Personal/Institutional Homepage
 - 1.2.4 Navigational Tools
 - Web Directories
 - Subject/Information Gateways
 - Search Engine
 - 1.2.5 Downloads
 - 1.2.6 Propaganda Materials
- 1.3 Multimedia Content
 - 1.3.1 Introduction
 - 1.3.2 Characteristics of Multimedia Applications
 - 1.3.3 Components of Multimedia Content
- 1.4 Summary
- 1.5 Answers to Self Check Exercises
- 1.6 Keywords
- 1.7 References and Further Reading

1.0 OBJECTIVES

For any information processing and retrieval system, analysis of contents is vital, as the effective analysis of contents will lead to organization of information, which should facilitate access to required information. It is equally true for Web based documents. It is necessary first that you know the different kinds of Web documents and their features.

After reading this unit you will be able to:

- get an overview of different types of web documents;
- get an idea of the need of these web documents; and
- learn about different kinds of multimedia file formats and their features.

1.1 INTRODUCTION

The inception of Internet has opened up new vistas of communication channels for information. It has generated a completely new kind of world, which is often referred to as Cyberspace, or in other words 'virtual world'. In this cyberspace web documents live as entities. But when we consider these entities as documents and

as documents and web as a collection of documents, the concept of a library prevails in the mind which has given rise to a new concept called a 'Virtual library'. Thus, often you will find that the Internet is referred to as a Virtual library particularly in the librarians' lobbies. Then question arises as to what are the documents this particular library has? As we know in libraries we can have books, journals, audio-video materials, images, and many other things, similarly on Internet you will find a variety of documents to fulfill different kinds of demands, including the ones we have described in the sense of our normal library. But Internet has all this in a different format i.e. electronic format as compared to traditional paper. There are different kinds of documents available on Internet like personal pages, e-books, e-journals besides a lot of propaganda material. The most important feature of current Internet is Multimedia. Multimedia as we know it is a combination of different media like a audio, video, image and text. It has emerged as the chief source of information as it creates a great impact on the human mind. Each has different kinds of file formats depending upon the degree of compactness and use.

1.2 TYPES OF DOCUMENTS

In the traditional environment we have different kinds of documents like books, journals, reference materials, images, audio-video materials etc. Besides, there are tools like catalogues or indexes available to assist in searching of information. Since we know that Internet is a virtual library, the items will be the same, only the format will change. Thus the varieties of documents available over Internet are:

- Books
- Journals/articles
- Personal/institutional home pages
- Search engines/directories/navigation tools
- Subject gateways
- Download repositories
- Propaganda materials
- Others like chat sites, music downloads, FAQ sites, art galleries, etc.

1.2.1 Books

In our day-to-day life books have a great importance. We use books from childhood till death in some form or the other. When we try to define a book it can be said that a book has the following features:

- It is a single (sometimes large) volume entity
- It has a cover page
- It has a title page
- It has pagination
- It contains a contents page
- It has lessons or chapters.

But the same is not true for Internet. E-books are common over Internet. They can be found in many formats, for example, Portable document format (PDF), PostScript format (PS), HTML (HyperText Markup Language) format. To read these formats e-book readers also are available. For PDF format Adobe Acrobat, for PS file Ghost View and for HTML document Browsers are available to read and these are easily downloadable.

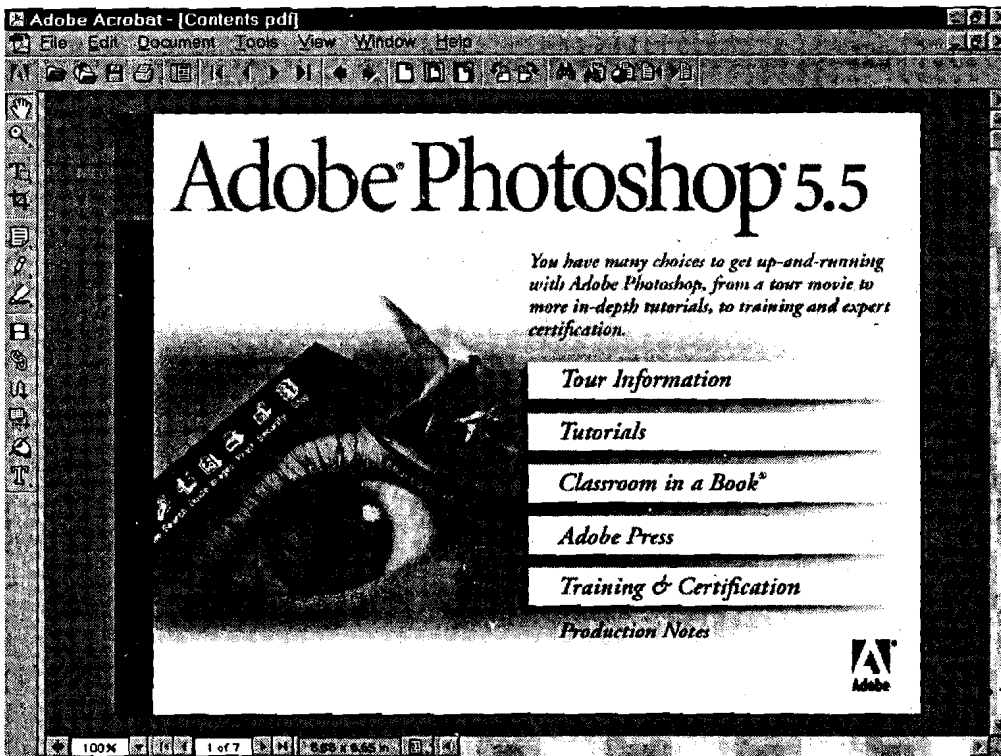


Fig.1: E-book in PDF format

Why books?

"Books are the quietest and most constant of friends; they are the most accessible and wisest of counsellors, and the most patient of teachers."

Charles W. Eliot, The Happy Life, 1896

Books in the true sense contain organized knowledge. When one wants to gain knowledge in a structured way he/she consults books. The basic organization of a book is from generic to specific. A book starts a subject from introduction and finally ends in very advanced topic.

1.2.2 Journals/Articles

Journals are publications, which appear periodically. Traditionally journals come in a booklet format, which contains articles. A journal also normally contains a content page and a cover page.

In the traditional libraries the most costly affair is the purchase of journals. Journals are usually very costly but contain a very high degree of intellectual material and nascent information. Online subscription of journals is becoming very popular now a days. Many of the libraries are going in for such subscriptions. On the Internet many of the journals are freely available but a majority are commercial, which means one needs to purchase the license for

use. In online subscription you can get access to the issue of journals as well as back volumes provided that you have asked for back issue access in your subscription. Since subscription of journals is a very costly affair the consortium approach is becoming popular day-by-day. In a consortium, a set of libraries subscribes to a journal and they acquire right to use among themselves. But there are many issues attached to online journal subscription. The first and foremost is speed of access and next is ownership, cost notwithstanding. Often it is found that the access speed is very slow. There are also hidden costs like payment to ISP, electricity etc.

These journals contain learned articles. Now-a-days lot of articles are available been through personal homepages. This particular approach has been generated because of the open source approach.

1.2.3 Personal/Institutional Homepage

A homepage is the index web page of an individual/institutional website. It may contain personal information about the author or topics. In case of an institute it gives information about the institute and its activity. It also includes links to the other pages of the site and maybe to other websites.

These are most common over Internet. Many people keep their personal information on web page. Besides, all the renowned institutes maintain their web pages.

To create a web page, HyperText MarkUp Language is mostly used. Thus in other words, "A homepage is an HTML document constructed in a similar way as a word processing document. This document is commonly referred to as a web page. A collection of these WebPages is what is commonly referred to as a website. A domain or dot.com is the collection of WebPages residing on the WWW with a name such as *www.yourname.com*."—Rick Thomas.)

Why website?

A web page may contain many basic elements; text, graphics, pictures, video, and sound. A web page is presented on a computer screen in a very similar fashion to that of a book. A page in a book may contain text, pictures, graphics, charts, etc. A web page can be a true multimedia exposition incorporating sound and video along with the text, graphics, pictures and charts.

For a library, a web page becomes a virtual electronic marketing tool. A library website can inform users about the services it offers. Websites can also contain a calendar of events, provide pictures, descriptions and pricing of the products and services that are offered by the business. A website may contain annual reports, frequently asked questions and answers about products and services, press releases and forms for feedback and membership forms with the ability to accept payments. A website also can convey to your users prospects and future plans of the library, offers, incentives, handle event registrations, provide surveys or questionnaires, contests etc.

Announcing Your Website

Making the website accessible on Internet is a key issue for libraries. Search

engines are one common solution to the problem. As we know, the search engines index the web pages and keep tabs in their databases. Popular search engines like Google (www.google.com), Yahoo (www.yahoo.com) and AltaVista (www.altavista.com) make a website accessible through search. One needs to type the keyword and the search engine performs the search.

Another approach is going on advertising yourself and Dr. Ranganathan opined “libraries should go to the users”. There are many ways a library can advertise itself. Often we encounter in the yellow pages about the different products where at the end of announcements the advertisement we see a URL with a caption “consult for further details or information”. Similarly libraries also put small advertisement in such yellow pages. Another approach is to advertise in the local newspapers where reference can be given to the library website.

A very common approach seen over the Internet is to advertise through others’ websites. Often one encounters advertisement about commercial products attached with the email. Similarly libraries can boost their website by advertising themselves through others websites, though netiquette should be followed and unsolicited mail should not be generated.

Benefits of a Homepage

With the public becoming more and more aware of time management, the Internet has truly taken its position in the daily lives and routines of people. Your library with a website now is able to make people know what you have in your library. From a business standpoint, it allows the business owners to streamline their communications with prospective customers, it pre-qualifies prospects and extends your marketing reach and lets you conduct business online. In fact a homepage gives the library a rightful place for users worldwide and also access to numerous Internet facilities.

Self Check Exercise

- 1) Describe the need of having a website for communication.

.....

1.2.4 Navigational Tools

To surf or search the Internet we require tools. These tools are known as navigational tools. Basically there are three kinds of navigational tools.

- Web Directories
- Subject Gateways
- Search engines

1.2.4.1 Web Directories

Web directories are a very common phenomenon over Internet. These are popular as they reduce the recall of search engine and increase the precision.

One can venture into the directory of any search engine and can experience increasing precision. The common example is seen searching for the term APPLE in search engines which also has a web directory. The search engine result brings the page related to APPLE as a fruit and APPLE computers. But the same search can be done using the directories of Agriculture and Computers separately. One can find now that whatever he is getting is only restricted to a particular directory. That means if the search is done in Agriculture directory the result output is only about the APPLE as fruit not APPLE computer. It implies that if someone needs more precision in his/her search he should look for something, which searches, in a particular subject. That is what the Web Directories are doing today. They act as catalogues or guides. Directories help not only in searches but also organize the information available on WWW and provide the links. Quite a good number of Web Directories are available over the Internet.

The organization of subjects is another issue. Mostly the web directories organize information in a hierarchical manner. The feature and organization vary from one directory to the other.

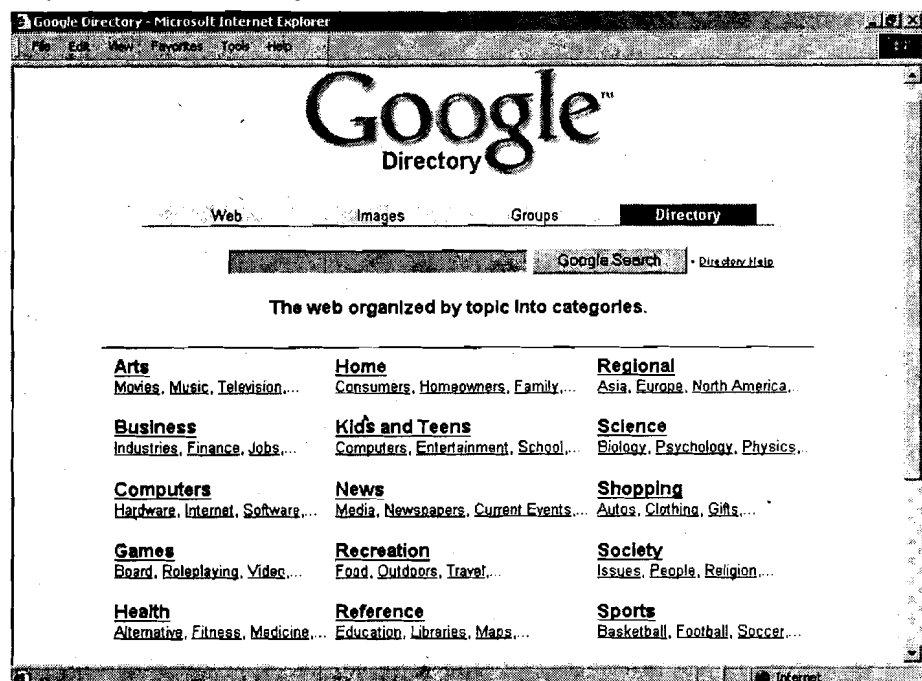


Fig. 2: Web directory of Google

Web directories use hypertext links to the information resources they list. These links facilitate immediate browsing of the information resources from the directory listings.

Why Web Directory?

Due to explosion of information over Internet it is felt that the results fetched by search engines are often mixed with unwanted ones. In order to increase the precision of search results it was felt that classifying the information in small subject directories is a good idea and then performing the search within the directory, which will reduce the recall of search result and thus increase the precision.

In a web directory information collected is reviewed, indexed, classified and organized on a web page as information resource, though this classification is

not as exhaustive as in a library. At the broader level it is very helpful to organize the resources.

Benefits of Web Directories

The web directories are basically indicators of information. In a broad sense, these directories are similar to an encyclopedia, where information about all the subjects is stored broadly. These directories link to subject specific search engines or to subject gateways.

1.2.4.2 Subject/Information Gateways

To provide quality information to the user community subject gateways are designed. They are valuable sources of information for users since they are subject specific in nature.

The biggest benefit of such sites is, they are man made. The information rendered is tailored manually by a person who knows the subject and its ins and outs. That is why the resource entries generated are usually superior to those available from a conventional search engine. (*IMesh Toolkit Project, UKOLN, ILRT at the University of Bristol*).

Information gateways are quality controlled information services that have the following characteristics:

- An online service that provides links to numerous other sites or documents on the Internet
- Selection of resources in an intellectual process according to published quality and scope criteria
- Intellectually produced content descriptions, in the spectrum between short annotation and review. A good but not necessary criterion is the existence of intellectually assigned keywords or controlled terms.
- Intellectually constructed browsing structure/classification
- At least partly, manually generated (bibliographic) metadata for the individual resources.

Why Subject/Information Gateway

With the advent of Internet many libraries are looking forward to going online with Internet. Often they find that the information available over Internet is enormous and if that alone could be some how managed and made accessible to the user community satisfy its information need. For the same they have devised subject-based portals, which are known as Subject Gateways in the information society. It is not necessary to a library to maintain a subject gateway. People who are information conscious and want to share useful information can manage it. The major idea of Subject gateways came with the inefficiency of search engines as they failed to give pinpointed information and because of information boom it is difficult for the academicians and researchers to sit and browse for information over Internet. Thus subject gateways proved to be a launch pad for the user who wants to get only useful information.

In the traditional library it was the librarian who used to work as information intermediary between the *right information and right user*. He was the one who filtered, catalogued and indexed the information so that users could easily

access information but on the web which is a virtual library who will do it for the users? Subject gateways are the answer, which work in anticipation but specific to subject domain. Thus to connect the users to right information these subject-based portals are very helpful sources.

They employ subject experts and information professionals to select, classify and catalogue Internet resources to aid search and retrieval for their users. Users are offered access to a database of Internet resource descriptions which they can search by keyword or browse by subject area. They can do this in the knowledge that they are looking at a “*quality controlled*” collection of resources. A description of each resource is provided to help users assess its origin, content and nature, enabling them to decide if it is worth investigating further.

1.2.4.3 Search Engine

A search engine fetches results from its database, which is developed by gathering information by “Web crawlers, spiders or robots”, which are nothing but programs which visit site by site and index the terms.

Search engines use keywords to locate information on the web. The results are presented in the order depending upon the algorithm followed by the search engine. There are search engines which give weightage to search results but quality is always doubtful because the weightage itself maybe subjective.

Why Search Engines?

Internet has become a collection of all kind of documents. But locating reliable information is the biggest challenge of the time. Search engines are one kind of solution to it. Almost all the search engines give combination search to increase the precision of search result.

There are search engine interfaces that search many search engines at a time. One such search engine is Askjeeves (www.askjeeves.com), which searches many search engines at a time. These metasearch engines search across many primary search engines and present the results together.

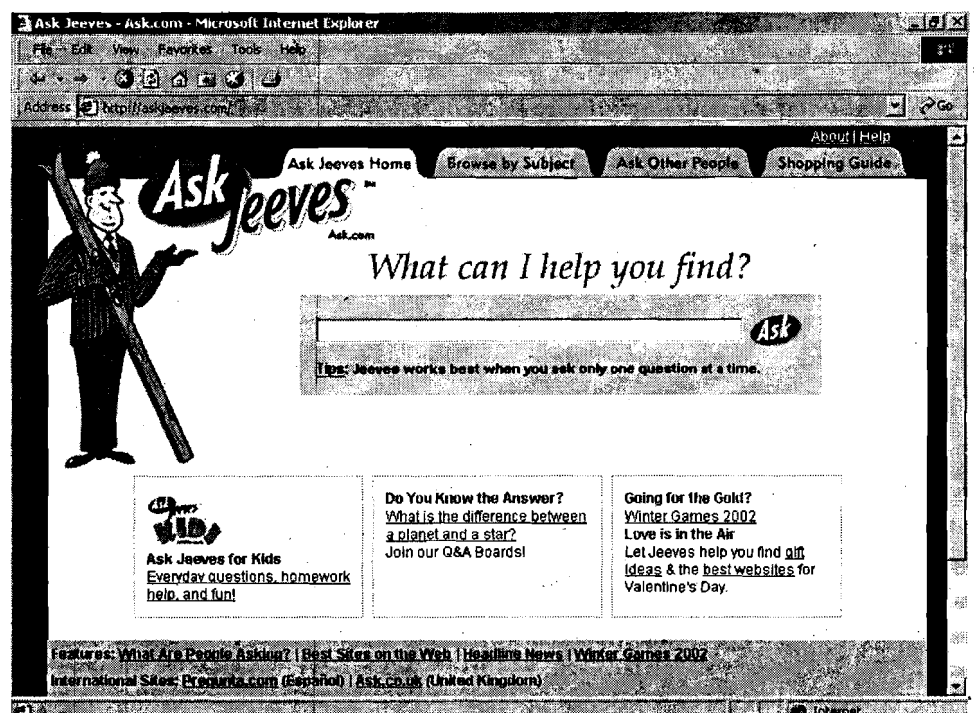


Fig. 3: Interface of Askjeeves

Self Check Exercise

2) What is a navigational tool? What are the navigational tools that are available over Internet?

.....

1.2.5 Downloads

Quite often we download files from Internet. These files are stored in repositories. On Internet many sites offer various software, which can be downloaded. The downloadable files can be software, device drivers data or even multimedia files. Normally they are stored in compressed format, so that they can be downloaded faster.

Now a days there are a number of software available for download management. These software can pick up any broken download. So one need not start download afresh in case of a power failure or a broken link. A few of such programs have the capability of enhancing the speed of download. Download accelerator is one such software which can be downloaded from www.speedbit.com [as of June, 2002]

1.2.6 Propaganda Materials

Often we see on Internet, a kind of information which is basically propaganda material. The literal meaning of it is ideas, facts, or allegations spread deliberately to further one's cause or to damage an opposing cause.

Often we get mails about some kind of product for which we have not at all applied or subscribed. It is a kind of push technology, which compels others to listen.

1.3 MULTIMEDIA CONTENT

1.3.1 Introduction

Information can be stored in any format, including text, image, audio and video. If the content of resource is in more than one medium but integrated into one it is called as a '*multimedia resource*'. Studies show that listening to something has more impact than reading, and watching something has more impact than listening. Since there is information in multimedia format it has become an obligation for libraries to store it and display whenever or wherever the user needs it.

When we look at the multimedia products we find that it is a combination of audio, video, text and image or atleast two of them. That means,

“Multimedia is the process of integrating different types of media into one format. In essence, it combines full motion video, pictures, sound recordings

and text into a single package that can be accessed by the user in a variety of ways. This process can be used for any type of application ranging from training systems, corporate presentation programs, interactive information kiosks and retail CD-ROM.”

1.3.2 Characteristics of Multimedia Applications

The popularity of multimedia is due to its many advantageous features.

Multimedia is **flexible, creative, powerful** and **cost effective**:

Flexible

- easy and inexpensive to update content
- utilization of very powerful programming tools
- playback on standard personal computers or other inexpensive systems
- presentations may be made to any size of group through the use of personal computers, overhead projectors or video projection systems.

Creative

- allows for the development of highly effective communication tools
- powerful graphic and video capabilities are available
- sophisticated new generation design tools are being developed
- 3-D graphics and animations are now possible.

Powerful

- more powerful and exciting than traditional presentation systems
- longer viewer attention spans
- higher information retention rates
- very effective teaching and communication medium.

Cost Effective

- due to rapid development in technology, prices of computer equipment are now gradually falling
- mass CD-ROM replication at a very low cost
- no specialized standalone equipment required
- can re-utilize existing content such as videos, brochures, text materials
- digital material can be easily updated without wholesale reproduction.

Self Check Exercise

3) What is multimedia? What are the characteristics of multimedia applications?

.....

.....

.....

.....

In general, we know that a multimedia application is a collection of audio, video, text and image. Each type has its characteristics and varied forms of storage formats.

Text

Libraries basically store books. These books contain text. Textual information is information written in a script which conveys meaning to those who read it. But computers do not understand any language or script. Computers understand everything in the terms of current on and off. There are several character encoding systems to represent the characters in computers for example, like EBCDIC (Extended Binary Coded Decimal Interchange Code), ASCII (American Standard Code for Information Interchange), UNICODE etc. The main character encoding system, which came in to existence and is still surviving, is ASCII. It was developed in 1968 and became the defacto standard. The major problem associated with ASCII is that it can't represent characters of other scripts because it is a 7-bit character code which can represent only 127 ($2^0+2^1+2^2+2^3+2^4+2^5+2^6=127$) characters at a time. But since there are 8-bits in a byte so ASCII can be extended to 256 ($2^0+2^1+2^2+2^3+2^4+2^5+2^6+2^7=256$) characters. But still this is not enough to accommodate more than 256 characters which need to be represented. For the same reason a new character encoding system was designed called UNICODE. UNICODE is a 16-bit character encoding system which can accommodate 65,536 characters at a time which is good enough to accommodate almost all the world scripts.

Basically two types of Text documents are found.

- Unformatted document
- Formatted document

Unformatted Document

There are many file formats in which we can store a text. Plain text files usually have the extension '.txt'. They are also called ASCII text files and can be viewed with an editor (such as Edit or Notepad) or with a Word Processor (such as MS Word or Word Perfect). The characteristic feature of plain '.txt' files is that they do not allow any kind of formatting on the document (such as bold, italics, font color, images, etc.). Many simple programs are written as text files. E-mails are also sent as plain ASCII text although now many mailing programs handle formatted mails.

.htm/.html files

These files are also text files written in text editor in which web pages are authored. 'HTML' stands for 'Hyper Text Markup Language'. The code of a web page is written in plain text and is saved with the extension '.htm/.html'. The browser (such as Netscape Navigator or Internet Explorer) identifies the file as a web page, reads the code and displays it on the screen as we see it with the images, colors and hyperlinks.

Formatted Documents

.doc , .wpd files

A very common format found on PCs, for formatted text files, '.doc' which

stands for 'document' files. These files may be created, viewed and edited using programs such as MS Word, Word Perfect and so on. Several formatting features such as bold, italics, justification, bulleting, to mention a few are possible. Now a days, e-mail programs also support formatted files.

.pdf files

'pdf' stands for 'Portable Document Format'. This file format was developed by Adobe Systems in order to make it possible to transfer formatted documents over the Net so that they look the same on any system. The biggest advantage of .pdf files is that it allows for printing of web pages - page by page as though it were a document file. This file type requires Adobe Acrobat Reader to view .pdf files and can be downloaded from the Net.

.ps files

'Post Script' files is also an ASCII file type that is technically plain text. However, it is unreadable unless an onscreen viewer like 'Ghostscript' is used. It can be read by a PS printer.

Image/Graphic Files

Images or graphic File formats are probably the most in number and variety among the four categories of file types. The reason may be due to the fact that images are one of the most important features of web pages, today. At the same time they have to be capable of downloading quickly, they cannot be bulky, though the resolution should be good. The several demands on graphic images have lead to the development of a variety of Graphic file formats - each with a distinct feature, advantage or disadvantage.

Before we discuss the various graphic file formats and their features, it may be useful to understand the basics of Graphics.

Graphic styles may be divided into two major types:

- Raster Graphics
- Vector Graphics

Raster Graphics

Raster Graphics/Images are simply a collection of dots known as 'pixels'. Hence they are also called 'bitmaps'. The color of each pixel is described by one or more information channels - separated into the primary hues - Red, Green and Blue or in a single stream of color mapped data. Raster images are simple images and are hence most suitable for interoperability. However, the primary disadvantage is that they do not scale well. Scaling may lead to a loss of resolution and hence poorer picture quality.

Vector graphics

The more complicated of the two, Vector Graphics define an image as a collection of vector equations. The obvious advantage of this type is smooth curves and lines irrespective of the size of the image or resolution. However the disadvantage is that they take longer to draw and require more storage space. Despite these disadvantages, vector graphics are gaining popularity.

Color is a very important feature of graphic images and different file formats pay a great deal of attention to the detail of color handling. The purpose of a project may define the extent of color clarity and quality that is required. For example, a website that primarily handles photographs requires at least 16-bit or near true color. However for most standard applications 8-bit or 256 colors is sufficient.

The following comparison explains the relation between bits and colors:

- 1 bit per pixel refers to an image with 2 colors
- 4 bits per pixel refers to an image with up to 16 colors
- 8 bits per pixel refers to an image with up to 256 colors
- 16 bits per pixel refers to an image with up to 32,768 colors
- 24 bits per pixel refers to an image with up to 16,777,216 colors (approx 16.7 million)

Using Graphics

Graphics are generally used in three ways:

- i) As viewed online or printed as a stand-alone document
- ii) Edited/Modified document using a graphic editor
- iii) Incorporated into a document that contains text, images, etc - Desk Top Publishing.

Desk Top Publishing includes text and graphics with emphasis on the presentation. This requires formatting. To save formatting, the text must be converted from one format to the other. Today, popular publishing include '.ps' and '.eps' and '.tiff' files.

The following sections discuss some of the most common graphic file formats in use today.

.bmp files

Bitmap files or .bmp files are the standard Windows Raster format. These files lay emphasis on quick display. Hence it stores images in the uncompressed form. The obvious trade off is that bmp files are space eaters.

Advantage:

- 1) Very quick download time
- 2) Supports 1,4,8,24 bits of color per pixel

Disadvantages:

- 1) Uncompressed - hence occupies more storage space
- 2) Transfer from other formats to bmp format is not handled well - loss of picture quality

.cgm files

Computer Graphics Metafile, is an ANSI standard graphic file format used to exchange vector data. It stores vector information as opposed to other file formats like gif, bmp, etc. that store raster information. As a result, vector files allow viewing with pan, zoom and without loss of detail. *cgm* files can be viewed directly on the web using a web browser plug-in.

Advantage:

- 1) Can be used as a format to transfer vector graphics such as post script files from Unix programs to Mac programs without a loss of resolution
- 2) Widely supported by Unix & Windows. Beginning to be supported by Macs.

Disadvantage:

- 1) Cannot be used with native browsers such as Netscape
- 2) *.cgm* files are very large (atleast 10 times larger than *.gif* files)

.gif files

One of the most popular graphic file formats on the Internet, Graphic Interchange Files (*.gif*) was developed by Compuserve with the main purpose of archiving information. *.gif* images are usually scanned stand-alone pictures that are not 'drawn' using an application program.

Advantage:

- 1) Highly compressible
- 2) Very useful format when large number of images are to be incorporated
- 3) A standard web format - most browsers have a *.gif* viewer
- 4) Small size allows quick transmission over the Net
- 5) Later versions of gif support transparent colors (which may be used in the background of webpages). Also support animations.

Disadvantage:

- 1) Not a good choice for non-web images
- 2) Supports only 256 colors - if true-color is desired, it is better to use *.tif* or *.jpeg*
- 3) If a higher resolution file is converted to the *.gif* format, the extra colors that *.gif* cannot handle are thrown away from the image, resulting in poorer quality.

.iff files

Amiga Interchange File Format (IFF) is an application specific format developed by Commodore Amiga for transfer to and from its own computers. This data format is designed for storage, exchange and manipulation of data between many different programs.

Advantages:

- 1) Image files can be easily processed by several programs sequentially, one after the other
- 2) When used with pictures, it preserves quality of color
- 3) .iff stores sound data, text and configuration data.

Disadvantages:

Can be used only on Amiga computers

.jpeg/.jpg files

JPEG stands for 'Joint Photographic Experts Group' that designed this format for high compression. It is one of the most popular image formats on the web. It discards extra data -or data beyond what the eye can see and hence has good compression capabilities.

Advantages:

- 1) It is a web standard (second only to .gif)
- 2) Image sizes are small - hence are quickly transferable over the Internet
- 3) Flexible - allows user to choose between image size and picture quality
- 4) Useful when large number of files are to be incorporated
- 5) Supports millions of colors
- 6) Picture quality remains accurate at lesser levels of compression
- 7) Usually better compression (compared to .gif) in photographs.

Disadvantages:

- 1) Very high levels of compression result in loss of picture quality
- 2) Repeated saving in this format results in loss of quality. If repeated saving is required, first save repeatedly in another format and save final image only in .jpg.

JAS files

JAS format is from JASC Inc. This file format is designed to create the smallest possible image file for 24 bits per pixel color images and 8 bits per pixel gray scale images. JAS results in "lossy" transformations: saving and retrieving an image using the JAS file format will result in some loss of image data. The amount of loss is dependent on the compression level that you have selected in your application. By using the lowest possible value for the file compression you will have the least amount of loss.

Advantages:

Very high level of compression

Disadvantages:

- 1) Supports only up to 24 bits
- 2) There is some loss of data during compression - poorer quality

.MAC files

An application specific format, MAC files are used in Macintosh MacPaint application.

Advantages:

- 1) Yields small file sizes.
- 2) It is still readable by most Mac machines.

Disadvantages:

- 1) Supports only 1-bit per pixel - only monochrome files can be converted to .MAC files.
- 2) .MAC format requires a fixed image size (576x720). Images converted to .MAC will be cropped to fit that size.

PBM plus files

PBM files are Portable Bitmap files, popular in UNIX

Advantages:

- 1) Allows for extensive manipulation of grayscale and color bitmap images.
- 2) Three libraries- pbm, pgm, ppm allow conversion of bitmaps to every other popular graphic file format.

PCX files

PCX file format was developed by Zsoft for its Paintbrush software. This format is used by IBM computers. Version 5 is most popular and supports 1,4,8,24 bits per pixel.

Advantages:

- 1) .pcx format is supported by more applications than any other format
- 2) Highly flexible
- 3) Accommodates any size and any number of colors
- 4) Image compression is an integral part of this format

Disadvantages:

Causes some difference in images if using different versions.

PICT files

This is the standard Apple Macintosh graphic file format

Advantages:

- 1) Accepted by many applications
- 2) May be imported/exported using clipboard (cut, copy, paste) to almost any text or graphics program.

Disadvantages:

Used only by Macs

RAS

Sun Microsystems developed the 'Raster' file format. There are three specific types of .RAS files

Type 0- Old style

Type 1- Modern style

Type 2 - Experimental style

Type 1 supports 1, 8, 24, 32 bits per pixel

Raw files

This flexible format consists of a stream of bytes that describes the color information in the file. Each pixel is described in binary format where 0=black and 255=white. This format is used to transfer documents between different applications.

TGA files

Targa files or TGA files were developed by Truevision. It is widely used by high-end paint programs and ray tracing packages.

Advantage:

- 1) Can handle images with upto 16 million unique colors.
- 2) May be saved in the compressed or uncompressed forms.

Disadvantage:

- 1) Not as widely used as .PCX or .TIFF files.
- 2) Was designed for use on systems that run MS-DOS color applications.

.tiff files

TIFF stands for Tagged Image File Format. This format was designed to overcome the problem of application dependence. It was originally designed to become the standard format. The format was intended to be capable of handling just about any possibility. This file format is generally used when graphic files need to be moved between different computer types (For example: PC to Mac and vice-versa).

Advantages:

- 1) Allows for high resolution
- 2) Highly flexible - there are several possibilities of how a .tif image can be saved
- 3) Is supported by most scanning and image editing software
- 4) Format works well for both on-screen display and print of photographs

- 5) Format works well in page-layout programs as it allows editing of file attributes
- 6) Can differentiate between types of images in three categories - Black & White, Gray scaled, Color.
- 7) Supports 24-bit true color.

Disadvantages:

- 1) .tif files from IBM PCs are usually different from those of Macs
- 2) The format is meant to be a standard. But many manufacturers have tried to 'improve' on it and it is no longer a standard
- 3) No application can claim to support all the variations of .tiff files
- 4) Files tend to be larger than many other formats and hence takes longer to print.

WPG files

Word Perfect Graphic files or WPG is an application specific file format. It first appeared with Word Perfect 5.0 and changed with Word Perfect 5.1.

Advantages:

Can contain bitmap, line art and vector images

Disadvantages:

- 1) If an application other than Word Perfect is used to view a WPG file containing bitmapped and vector elements, vector elements will be discarded
- 2) Supports only up to 8-bits of color per pixel.

Audio File Formats

Sound files or audio files are gaining popularity on the web. Today, most of the latest sound tracks are available on the Internet as sound files. There are even a few albums that have their presence on the Web alone. Another popular application is online live news broadcasting. The following section discusses some of the common audio file formats:

.au files

Most commonly found on the web, it is required by PC users to load applications such as Waveform Hold and Modify to play these files. Macs need different sound applications to play this file type.

.midi files

This is used by files following the Musical Instrument Digital Interface standard. These are used mostly in audio control in Multimedia industry. MIDI file specification allows for lengths to be specified as a variable number of bytes.

.aiff files

Audio Interchange File Format (aiff) was developed by Apple. Although it

was originally made for Macs, now it can be used by other platforms too. It is a very good audio file format for use on the Internet. It can also be used in Multimedia authoring on both Macs and Windows.

.mp3

mp3 -stands for MPEG layer three. It is currently the most popular of the audio file formats. Its hallmark is its CD-quality of music. MP3 allows for very high levels of compression. A minute of music may constitute approximately 1 Mb file. An MP3 player that is readily available on both Macs and Windows is required to play this file type.

.voc files

Creative Lab's Sound Blaster uses the .VOC file. They are designed for storing digitized voice data and hence the name. They can however also handle any digitized sound in any of a variety of formats. The VOC files have a two part structure. The header block which defines the contents of the file, the data block which actually contains the audio information.

.wav

Wave files is a commonly used file format on Windows machines. It can be used on the Internet and is good for multimedia authoring. It is flexible and handles both compressed and uncompressed storage formats.

Video Files Formats

Video files have become most popular with films being available and viewed on VCDs and DVDs. However within a multimedia lab, it is important to be aware of the video file formats are these as the most disk space-occupying (bulky) types.

.avi files

Audio-Video Interleaved file format was developed by Microsoft. An AVI player and drivers are required to play this format. They are readily available both in Mac and Windows machines. With the player, AVI plays a full motion picture video with audio in a small window at about 15 frames per second.

Advantages:

- 1) AVI comes with Windows, so no drivers need to be obtained and has the built in media player, although the Indeo drivers are better for faster machines and will improve quality.
- 2) AVI is a popular standard, many videos have been produced in the format because of it's non requirement of drivers.
- 3) The quality of AVI files with good drivers and good hardware can be quite impressive.
- 4) The majority of AVI files have audio.

Disadvantages:

- 1) AVI's are no longer developed by Microsoft and are left to be developed by a third party.

- 2) Frame rates are not as high as other video formats (such as
- 3) MPEGs)
- 4) Not much development is going on in this file format.

.mov/.movie files

Movie files are the common format used in QuickTime movies, the Mac native video platform.

.mpg/.mpeg files

The standard Internet format uses MPEG compression scheme. This format can be used on Macs by converting into QuickTime movies using applications such as 'Sparkle'.

.qt files

QuickTime files. The latest version is used on Macs today.

.ram files

Developed by Real networks for streaming video. This requires Real Player for viewing.

Plug-ins:

Often we want to add new features to our software. For the same we require to add new programs. These programs get attached with the existing software and add to the functionality of software. These small programs are known as plug-ins.

“Thus, a software module that adds a specific feature or service to a larger system is known as Plug-ins.”

For example, there are a number of plug-ins for the Netscape Navigator browser that enable it to display different types of audio or video messages. Navigator plug-ins are based on MIME file types.

1.4 SUMMARY

In this unit we have discussed:

The nature of content in different kinds of documents traditional as well as electronic. Internet is full of content and it has different kinds of documents. These documents range from books, journals to downloadable programs.

The most common type of documents available over Internet are Institutional/ Personal web pages, which provides individual or institutional information.

To search such pages over Internet there are a variety of tools like Web directories, subject gateways and search engines. Search engines are the most common tools for Internet searching.

Internet is full of multimedia contents. A multimedia format is a single file

having a combination of different kinds of formats like audio, video, text and image.

Audio, video, text and image, all these kinds of contents can be stored in different kinds of file formats. These file formats have advantages and disadvantages depending on the degree of compression and characteristics.

1.5 ANSWERS TO SELF CHECK EXERCISES

- 1) A website is very strong communication channel for any organization to propagate its product. The biggest advantage with a website is that once it is launched it is available round the clock and provides information about your products to people.

Libraries can also use it as a virtual electronic marketing tool. A library website can inform users about the services it offers. Websites can also contain a calendar of events, provide pictures, descriptions and pricing of the products and services that are offered by the business. A website can also contain annual reports, frequently asked questions and answers about your products and services, press releases and forms for feedback and membership forms with the ability to accept payments. In a library environment a website also can convey to your users prospects and future plans of your library, offers incentives, handle event registrations, provide surveys or questionnaires, contests etc.

- 2) Navigational tools are aids for searching any information available over the Internet. There are three kinds of navigational tools available:

Web directories: A web directory is a navigational tool. Often it is referred to as a Catalog also. These are simple listing of the websites. Basically a web directory is an indicator to information about where a link to the information is stored. The arrangement or structure of such directories is not fixed. These websites contain information on a variety of subjects. These sites can be considered as a launch pad to start the search.

Subject gateways: Often researchers and scholars find themselves overloaded with information. To help them in finding value added and quality controlled information the idea of subject gateways came into existence. A subject/information gateway is a web site that provides searchable and browsable access to online resources focused around a specific subject. Subject gateway resource descriptions are usually created manually rather than being generated via an automated process. Because the resource entries are generated by hand they are usually superior to those available from a conventional web search engine.

Search engines: These are the most popular navigational tools. Here one can use a search term and can get the search results. Behind the screen these search engines use robots or crawlers to collect information from the web and store that information in their database and whenever a query is fired the search is carried out on the stored database. The main problem associated with these search engines is that they retrieve a lot of irrelevant information.

- 3) Multimedia is a collection of different media like audio, video, text and movie or their combination in a single format.

Multimedia applications have the following characteristics:

Flexible : Multimedia tools are basically flexible. They can be replayed and recorded whenever and wherever it is necessary. To prepare a multimedia presentation one needs to have a very powerful computer which can support high resolution, audio and video.

Creative : It allows for the development of highly effective communication tools. A multimedia application can have even 3-D graphics and animations.

Powerful : It is more powerful and exciting than traditional presentation systems. It has viewer attention and deep impact.

Cost Effective : Costwise, making a multimedia application can't be considered as cheap. But it has a powerful impact on the audience.

1.6 KEYWORDS

Byte	:	It is unit of memory. A byte contains 8-bits.
Directory	:	It is a storage area.
Homepage	:	The first page of the website.
Multimedia	:	Combination of various media like text, images, animation, etc.
Navigation	:	Searching or browsing something with an aid.
Propaganda material	:	Noisy material. Generally, material published to further one's cause or to publicize a product.
Website	:	An electronic document, which is used to host information on the Web.

1.7 REFERENCES AND FURTHER READING

DESIRE Information Gateway Handbook.

<http://www.desire.org/handbook/contents.html>

What is multimedia? (<http://www.blackstone.ca/blackstone/whatis.asp>)

Download file formats and extension.

<http://www.learnthenet.com/english/html/34filext.htm>

Frater, Harald and Paulissen, Dirk (1994). Multimedia Mania. Abacus: Grand Rapids, 1994.

Graphics File Formats. <http://www.why-not.com/articles/formats.htm>

Perry, Paul (1994). Multimedia Developers Guide. Sams Publishing : Indianapolis, 1994

Review of Internet file formats. http://dio.cdlr.strath.ac.uk/file_formats.html

Rimmer, Steve (1994). *Converting and using Graphic files*. Sams Publishing : Indianapolis. 1994.

Rosch, Winn L (1995). *Multimedia Bible*. Sams Publishing, Indianapolis, 1995.

Thomas, Rick. What is a homepage anyway? *The Business Journal*, 8(5).

<http://www.thepbj.com/030300/a12.htm>

What are the most common file formats?

<http://www.public.asu.edu/~eve123/WhatAreTheMostCommonFileFormats.html>

What is a Web Directory.

http://www.lib.utc.edu/pages/web5_directories/sld02.html

Wodaski, Ron (1994). *Multimedia Madness*. Sams Publishing : Indianapolis. 1994.