
Chapter 18. The Internet as a Digital Library

Table of Contents

Introduction to the Internet as a Digital Library	1
Context	1
Introduction	2
Objectives	2
Conventional Libraries, Digital Libraries and the World Wide Web	2
The World Wide Web	3
Storing the information	3
Accessing the information	4
Interaction between client and server	5
Search Engines	6
Information Networking	6
Activities>	7
Activity 1 -	7
Activity 2 - HTML	7
Activity 3 - Data Network	7
Activity 4 - HTML 2	7
Activity 5 - Library Organisation	8
Activity 6 - Yahoo and Wikipedia	8
Activity 7 - World Wide Web	8
Review Questions	8
Review Question 1	8
Review Question 2	8
Review Question 3	8
Review Question 4	8
Review Question 5	8
Discussion Topics	9
Answers and Comments	9
Activity 1	9
Activity 3	9
Activity 4	9
Activity 5	9
Activity 7	10
Review Question 1	10
Review Question 2	10
Review Question 3	10
Review Question 5	10

Introduction to the Internet as a Digital Library

Context

This chapter builds on the previous chapters that have dealt with the structure and operation of the Internet and its software, and also with its application software.

Introduction

The use of the as World Wide Web a digital library is explained, along with its operation. As a digital library, it can be seen as a collection of digital objects and, equally, as the networked equivalent of a real-world library. The World Wide Web is a distributed system and, from this, the importance of storing the material in a standard way and the need to access it in a standard way can be explained. This allows the roles of HTTP and HTML to be made clear, in that one is its basic form of communication and the other the way of standardising the form of the material held by the servers. In addition, the fundamental importance of the client-server mode of communication is stressed.

The World Wide Web is a multi-media system, transporting data and meta-data to give meaning to the data. It is shown that the World Wide Web, along with enhancements such as search engines, provides a digital library capable of meeting the general needs of information workers.

Objectives

At the end of this module, you should be able to:

- describe the operation of the World Wide Web, including the contributions of HTTP and HTML.;
- appreciate the role of multi-media in the World Wide Web;
- establish that the World Wide Web may, with the support available, be regarded as a multi-media library capable of meeting the requirements of information workers.

Conventional Libraries, Digital Libraries and the World Wide Web

In parallel with this unit, you should read also read relevant chapters from your textbooks.

A conventional library is a more or less coherent and comprehensive collection of information: it so happens that the information is stored on paper. As a unified and comprehensive collection of information, the World Wide Web can also be seen as a library: the fact that the information is stored electronically in a digital format leads us to refer to it as a digital library.

As a digital library, the World Wide Web differs from conventional libraries in certain ways, some of which offer improvements, or the potential for improvement, while some do not. It can provide multimedia material, which a paper-based library cannot. It can be accessed from anywhere in the world, while a conventional library is inevitably for local access. Anyone with access to the Web can not only read the material it holds but also, with a little expertise, publish their own material. This has its good and not-so-good sides. Allowing anyone who wants to publish to do so banishes the assurance of reliability, authority and taste that are associated with the information to be found in a conventional library. Of course, reliable and authoritative material is published on the Web, but readers must evaluate what they find for themselves.

With a digital library, as with a conventional one, information must be stored and must then be made accessible. On the Web, the information is stored in documents on Internet computers, which act as Web servers, and is made available from them on request. Before they are stored, the documents are marked up using HTML. A document is accessed by using a browser to send a request for it and, when it is received, to display it in accordance with the way it is marked up. The interchange takes place using the communication protocol known as HTTP.

Ad-hoc browsing through individual web-sites is not an efficient way to locate specific information. Search engines are available to index the contents of the World Wide Web and provide for efficient search. In this way, the World Wide Web, as an accessible, organised repository of information, becomes a digital library.

To Do

Carry out Activity 1.

The World Wide Web

We now turn to the World Wide Web to consider how it stores information and allows for its access. Since storing information and providing for access to it are the basic ingredients for any library, we can see how the Web performs in these areas, and whether, as a digital library, it suffers from any relative shortcoming or possesses any natural advantages.

Storing the information

The documents stored on a Web server are structured, broadly speaking, into 'pages', with a page corresponding to a display screen. The documents are created using HTML (hypertext mark-up language) which marks up the text and other content of document in such a way as to describe the appearance it should have when displayed. Special marks are used to indicate, for example, that certain text is the heading of a document or the body of a document. Other marks can be placed before and after a section of text to indicate that it is to be displayed in bold or in italics.

One document can be linked to another, which can be on the same server or on a different one. The link is made by giving the location of the linked document, that is its address. The address consists of the address of the computer and the location in that computer's file store of the file containing the document.

In this way, the contents of the World Wide Web are stored on the numerous Web servers scattered across the Internet, but those contents are linked together to form a single 'web'.

The following simple example illustrates the form of an HTML document:

```
<html>

  <head>

    <TITLE> Simple Example</TITLE>

  </head>

  <body>

    <H1>Heading</H1>

    <P>A short illustrative paragraph that includes a single link to <A HREF=

  </body>

</html>
```

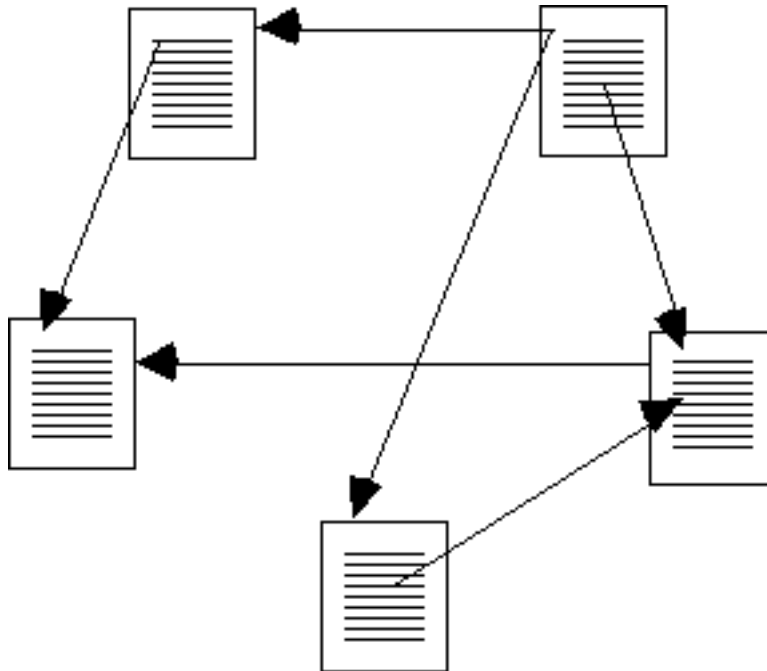
The HTML code above will be interpreted and displayed by a browser in a form similar to that shown below:

Simple Example

1. Heading

A short illustrative paragraph that includes a single link to Internet.

The ability to create links allows the content of the Web to be linked, and so to form what is essentially a single database regardless of the fact that it is distributed and, in fact, distributed world-wide. The resulting 'data network' can be visualised in the following way:



To Do

Do Review Question 1.

Accessing the information

Requests for information are sent from a browser. The request gives the location of the required item in the same way as a link does. A request has the form:

`http://<computer name>/<file path name>`

where <computer name> denotes the name of the computer acting as a Web server that holds the required file, and <file path name> denotes the path name for the file on that computer. A typical request could be:

`http://www.mdx.ac.uk/courses/bit.html`

When the requested document is returned, the browser, after interpreting the HTML marking, displays it in the appropriate fashion

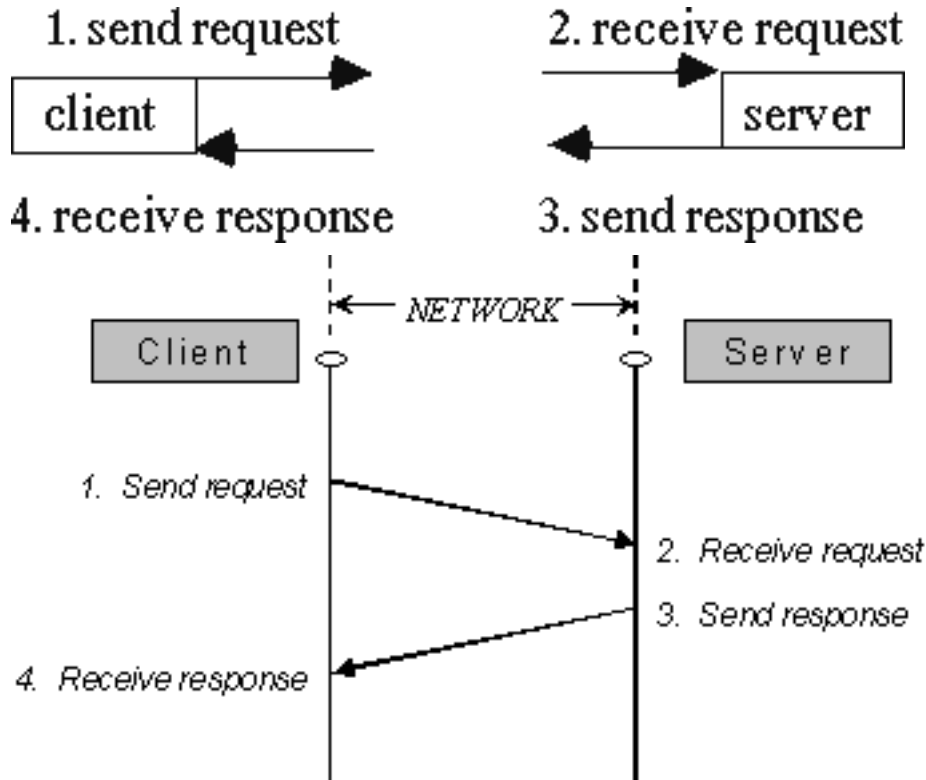
To Do

Do Review Question 2.

Carry out Activities 2, 3 and 4.

Interaction between client and server

The generic client-server interaction proceeds through four stages as follows:



On the Web, an ordinary computer such as a PC becomes a 'client' by running a browser; any Web server is a 'server'; and the name of the client-server interaction process is http.

The stages of the client-server interaction proceed as follows:

1. *Send request.*

From the client, which is running a browser, a request is entered by typing, for example, 'http://www.mdx.ac.uk/courses/bit.html'. This is taken as a request to the computer named www.mdx.ac.uk for the file '/courses/bit.html'. The request is sent to that computer for that file.

2. *Receive request.*

The named computer (a Web server) receives the request and locates the file in its file store.

3. *Send response.*

The Web server sends the requested file to the requesting computer. The file is an HTML file, beginning with <html> and ending with </html>.

4. *Receive response.*

The client machine receives the file and its browser displays it in accordance with the way it is marked up.

This is the essence of the interaction carried out by HTTP. Before this, HTTP establishes a path from the client to the server and afterwards it releases the path.

To Do

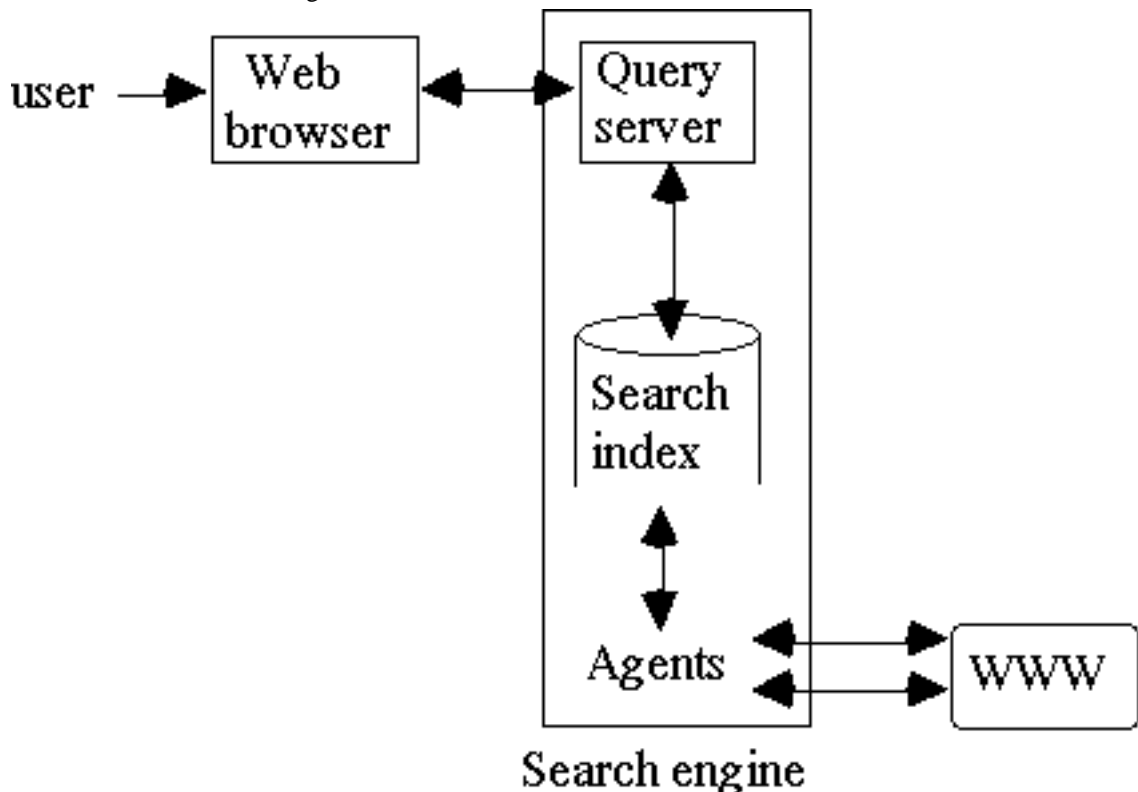
Do Review Question 3.

Carry out Activity 5.

Search Engines

The basic way to use the World Wide Web is to enter an address or to click on a link. This is fine if you know the address of the server that holds what you want, or if you can find a link that leads to what you want, but it is not so good if you are new to the Web or want to find information when you have no links to it. Search engines are intended to overcome the latter problems by indexing the content of the Web so that, to the user of the search engine, the World Wide Web seems more like a database.

The structure of a search engine is:



The agents search the World Wide Web for documents and bring back their details to be indexed. (Agents were dealt with in unit 14.) This means that the index is continually updated as new documents are added to the Web. Users interact with the search index and need not be aware of how it is created. The search index is, in effect, a database and can be queried using database queries. By formulating requests accurately and then narrowing or broadening them as necessary, information can be located precisely. As mentioned above, it then remains to assess the reliability and authority of what has been found.

To Do

Do Review Question 4.

Carry out Activity 6.

Information Networking

Information networking can be defined as the use of a networked computer to meet all the needs of an information worker. Of course, if the network can act as a digital library, the chances of its successfully supporting information networking are enhanced immeasurably.

The World Wide Web, in its guise as a digital library, and with the added capabilities of search engines, provides much of what is needed to carry out the information-related tasks that comprise information networking.

- It may be that it provides all that is needed.
- It is clear that information can be located, collected and aggregated in ways analogous to those that are used with a conventional library but that these activities are enhanced by the application of computer and, in particular, database technology.
- The involvement of databases supports the storage and rapid searching of large amounts of information.
- The use of computers allows software-based tools for text manipulation and information organisation to be brought to bear.

To Do

Carry out Activity 7.

Activities>

Activity 1 -

Make a systematic comparison of conventional and digital libraries by filling in the slots in the following table:

	Conventional Library	Digital Library
Information recorded on		
Media		
Access for readers		
Access for writers		
How to find items		
Nature of content		

You can find a discussion of this activity at the end of the chapter.

Activity 2 - HTML

The treatment of HTML presented here is deliberately brief, because extended treatments are provided by other modules (previously studied or to follow). If you have studied HTML it already, you should revise what you know. If you have not, note that it will be covered in another module.

Activity 3 - Data Network

The ability to link documents, whether they are stored on the same computer or on different ones allows for the creation of a 'data network' consisting of linked items of information. Describe the relationship between this 'data network' and the computer network on which it is supported.

You can find a discussion of this activity at the end of the chapter.

Activity 4 - HTML 2

One way to create an HTML document is to type the text of the document and then to mark it up manually. Another is to type the text into a word processor and then to invoke 'save as HTML'. Compare these two ways of proceeding.

You can find a discussion of this activity at the end of the chapter.

Activity 5 - Library Organisation

Describe the way in which items in a conventional library are dealt with to make possible to organise and present them in a systematic fashion. What is the nearest approach to this on the Web?

You can find a discussion of this activity at the end of the chapter.

Activity 6 - Yahoo and Wikipedia

Find out how Yahoo [<http://www.yahoo.com>] and Wikipedia [<http://en.wikipedia.org>]. Contrast and compare the two services.

Activity 7 - World Wide Web

Information workers would expect their computers, when connected to properly resourced computer networks, to meet all their information needs. These needs include, among others, the location, retrieval, refinement and integration of information.

Describe and discuss the extent to which the World Wide Web provides a complete environment for information networking.

You can find a discussion of this activity at the end of the chapter.

Review Questions

Review Question 1

In the context of the Word-Wide Web, what is a link?

You can find an answer/comment for this review question at the end of the chapter.

Review Question 2

How can *http://<computer name>/<file name>* be expressed in plain English as an equivalent request?

You can find an answer/comment for this review question at the end of the chapter.

Review Question 3

At this stage, with standard means of storing information and accessing it, how does the Web compare with a conventional library with the same capabilities?

You can find an answer/comment for this review question at the end of the chapter.

Review Question 4

What are the names of some of the major search engines? In what way are they different?

Review Question 5

At this stage, with standard means of storing information and accessing it, and with search engines, how does the Web compare with a conventional library the contents of which have been classified and indexed?

You can find an answer/comment for this review question at the end of the chapter.

Discussion Topics

1. Another approach to bringing the Web under control is to look for specific structures within its interlinked data space. For example, a strongly connected ‘island’ of pages that is loosely connected to the rest of the Web could indicate a body of knowledge restricted to a single topic. The truth or otherwise of this suspicion could easily be tested.

Which other structures would prove useful? How could structures be automatically detected? How could their value be automatically assessed?

2. Would a ‘map’ of the local area of the Web’s data space assist an information worker?

What would a ‘map’ be like? What does ‘local’ mean in this context? How would a map help? (Or, why wouldn’t it?)

3. Would it help information workers if there were two distinct Internets - one for ‘serious’ activities such as those of academics and researchers, and one for ‘recreational’ usage?

Answers and Comments

Activity 1

	Conventional Library	Digital Library
Information recorded on	paper	electronic storage
Media	text, images	multimedia
Access for readers	at its location during opening hours	any place, any time
Access for writers	published authors	anyone
How to find items	index	search engine
Nature of content	reliable, authoritative	diverse. Not necessarily reliable

Activity 3

Each item is stored on a computer. A link between two items stored on the same computer corresponds to a pointer from one place in that computer’s memory to another. A link between two items stored on different computers will not, in general, correspond to a physical link since a pair of computers chosen at random are unlikely to be directly connected to each other. Such a link will, however, correspond to a route in the network.

Activity 4

To mark up a text manually, it is obviously essential to learn the mark up language and then to take the time to mark up the text. If you can do this, and you can imagine the way you want the finished document to look, it is possible (within the limitations of HTML) to get exactly what you want and to get it with an efficient use of markings. On the other hand, a decent word processor shows you what the document looks like as you type the text, generates the marked up version automatically and saves you the trouble of learning HTML.

Activity 5

In a conventional library, items are classified according to their content using a standard scheme such as Universal Decimal Classification. Items can then be identified and located, either individually or according to their class, by their classification number.

The Web has no overall classification scheme. The nearest approach to it is a search engine, which allows items to be located by their title or by the use of key words.

Activity 7

Clearly, the World Wide Web supports the location and retrieval of information. It also contains information on most topics.

A computer has to be used to access the Web, of course, and that computer can be used to run the word processors, databases and so on that can be used to refine and consolidate the information obtained in one way or another from the Web.

Review Question 1

A link is a reference to another document, or a place in a document. As such, it consists of the name of the file containing the document (which can be a complete path name) or a reference to a location within the file containing the document.

Review Question 2

“Get me the file <file name> from the computer <computer name>.”

Review Question 3

Items can be placed in either type of library in more or less equivalent ways. With the Web, a specified item can be located more easily than in a conventional library that has not yet employed some means of organising its contents. Neither type of library allows items other than specific, known ones to be located at all effectively.

Review Question 5

One could claim that it compares rather well, in that it gives faster location and retrieval of items from a much larger selection.