Title
Unit 148 - World Wide Web Basics

Permalink
https://escholarship.org/uc/item/21n676tv

Authors
148, CC in GIScience
Yeung, Albert K.

Publication Date
2000

Peer reviewed
Unit 148 - World Wide Web Basics

By Albert K. Yeung
Ontario Ministry of Northern Development and Mines, Canada

This unit is part of the NCGIA Core Curriculum in Geographic Information Science. These materials may be used for study, research, and education, but please credit the author, Albert K. Yeung, and the project, NCGIA Core Curriculum in GIScience. All commercial rights reserved. Copyright 1999 by Albert K. Yeung.

Advanced Organizer

Topics covered in this unit

This unit explains the characteristics and the working principles of the World Wide Web as the most important protocol of the Internet. Topics covered in this unit include:

- characteristics of the World Wide Web
- using the World Wide Web for the dissemination of information on the Internet
- using the World Wide Web for the retrieval of information from the Internet

Learning Outcomes

After learning the materials covered in this unit, students should be able to:

- define terminology pertaining to the protocol and language of the World Wide Web
- describe the method of posting information on the World Wide Web
- identify client applications to retrieve information from the World Wide Web

Full Table of Contents

Metadata and Revision History
World Wide Web Basics

1. What is the World Wide Web?

- the World Wide Web is also commonly referred to as the Web, WWW and W3

1.1. The World Wide Web and the Internet

- the World Wide Web is one of the protocols of the Internet
  - the Internet is a super network made up of regional, national and international telecommunication networks linking together computers in educational institutions, government departments, military establishments as well as commercial and non-commercial organizations all over the world
  - users of the Internet communicate by the following protocols:
    - *World Wide Web*, a hypermedia system for delivering digital files in multiple forms, including text, picture, sound and animation
    - *file transfer protocol* (ftp), for exchanging files between computers
    - *telnet*, a log-on procedure for accessing programs on remote computers as though they were local
    - *electronic mail* (e-mail) a mailing system whereby messages among Internet users, and among Internet users and users of networks outside the Internet, are delivered and exchanged
    - *gopher*, a communication protocol for retrieving text- and file-based information distributed in different computers across the Internet
    - *newsgroups*, discussion groups which contribute information to communities of users in a particular area of interest
    - *wide area information searching* (WAIS), a distributed document retrieval system
  - the World Wide Web has practically subsumed many of the functions of the above protocols because
    - it is interactive and dynamic, allowing information to be updated easily
    - it is cross-platform (i.e. can be used on Windows- and Macintosh-based PCs and UNIX-based workstations)
    - it has a standardized hypermedia format that makes navigation in the Internet quick, intuitive and consistent
    - it is distributed globally across tens of thousands of sites, providing instant access to an impressive magnitude and depth of information in all subject areas
  - the World Wide Web is now the *de facto* graphical interface of the Internet

1.2. Who runs the World Wide Web?

- the Web was proposed in 1989 by Tim Berners-Lee of the European Laboratory for Particle Physics (CERN) in Switzerland to facilitate information sharing and dissemination to support groups around the world
for several years, CERN was the center of the development of the Web
- in late 1995, CERN passed its part in the Web to the Institut Nationale pour la Recherche en Informatique et Automatique (INRIA) in France
- at present, the World Wide Web Consortium (W3C) is the organization that coordinates development activities of the Web
  - it is based at the Massachusetts Institute of Technology (MIT) in the United States, INRIA in France and Keio University in Japan
  - it consists of individuals and organizations interested in defining and supporting the architecture, interface and languages, and technologies that make up the Web
  - its services include: repository of information about the Web for developers and users, reference code implementations to embody and promote standards, prototypes and sample applications to demonstrate the use of new technologies
- development of the Web is also strongly influenced by commercial browser vendors
  - a Web browser is a computer program that serves as the interface between a user and the Web (see 1.4.5)
  - the two major commercial browser vendors are Netscape Communications Corporation and Microsoft Corporation
    - Netscape is the vendor of the Netscape Navigator browser
    - Microsoft is the vendor of the Internet Explorer browser
  - although both vendors claim to support and adhere to the guidelines of the W3 Consortium, they also include their own proprietary features
    - features in Netscape Navigator and Internet Explorer often conflict with one another and with the work of the W3 Consortium

1.3. What is available on the Web?

- materials in the Web are stored digitally and can be in the forms of text, graphics, audio and video
- these materials cover an enormous amount of information in practically all fields of study and subject areas that one can possibly think of
  - the greatest sources of information are from universities, colleges and research institutions
    - these include research proposals and reports, theses, dissertations, course notes and other forms of instructional materials, book reviews, conference proceedings
    - increasingly, university and public libraries are making their information holdings available in the Web (see 3 and 4)
  - different levels of government have developed Web sites to provide the public with information about community and social services, economic development, the environment, natural resources, land management and land use, surveying and mapping, taxation, licenses and transportation
  - many commercial organizations have taken advantage of the Web to provide potential clients with information about their respective goods and services, research and development reports and, increasingly, facilities for electronic commerce
  - non-commercial organizations (e.g. the World Bank, the Organization for Economic Cooperation and Development, the World Wide Web Consortium and
the Open GIS Consortium, among numerous others) have also made use of the Web to distribute information about their activities, services and research reports
• numerous *individuals* now use the Web to disseminate information ranging from personal opinions to large collections of Web resources in specific subject areas

1.4. Concepts and Definitions of terminology pertaining to the Web

1.4.1. Client/server computing

• the Web operates in the client/server model of computing (Figure 1)
  • a *client* is a computer which requests computing services from another computer, known as a *server*
  • the client can be any computer but the server is always a computer that is specially devoted to providing services to client computers
• in the client/server architecture, a client can access many servers and a server can have many clients
  • this means that on the Web, a client computer can request information from different server computers at the same time
  • similarly, a server computer can provide information to different client computers simultaneously
• there are two approaches to dividing the work between the client and the server
  • the "fat server" or "thin client" approach places most of the processing functions on the server
  • the "fat client" approach places most of the processing functions on the client
• client/server processes are normally initiated by a user input from the client computer, and the server responds only when it receives a request from the client
• it is possible to automate the transmission of data from the server to selected clients using the "client pull" and "server push" mechanisms
  • "client pull" uses a directive to instruct the browser to reload or update a document from the server at regular time intervals
  • "server push" continually sends data to selected clients at specified time intervals
    • this process usually continues for an indefinite period of time, until the server knows it is done sending data to the clients, or when the clients interrupt the process

1.4.2. HyperText Transmission Protocol (HTTP)

• *HyperText Transmission Protocol* (HTTP) is the language by which computers in the Internet communicate with one another
• connections between computers using this language are described as the "stateless" or "query-response" model of interaction
  • there is no permanent connection between the client and server computers
  • after a request has been made by a client, the connection is broken
  • then a response from the server is sent back, and the connection is broken again
  • this process is repeated for every request and often even for a part of a request

1.4.3. HyperText Markup Language (HTML)
HyperText Markup Language (HTML) is the language used to write Web documents

- it is a language, but not in the sense of computer programming language such as FORTRAN, C and Visual Basic
- it is aptly called a "language" because, just like natural human languages, it contains all the rules (grammar) and codes (words and phrases) necessary for the creation of a usable document
- it uses standard ASCII characters and contains formatting codes, called commands or tags (Figure 2), that describe the structure of a document, provide font and graphics information and contain hyperlinks to other Web pages and Internet resources

1.4.4. Hypertext and hypermedia

- hypertext is the method by which hyperlinks (also called hotlinks) are incorporated in an HTML document that makes it possible to seamlessly refer to other HTML documents and retrieve information from them
  - when text, image, audio and video files are linked together by means of hyperlinks, the term hypermedia is used instead
- by using hyperlinks, the Web allows a logical connection of files, in much the same way as the human brain links associated pieces of information with one another (Figure 3)
  - a hyperlink is made up of two parts: an anchor and a pointer
    - the anchor appears on the computer screen as one or more underlined words
    - each anchor has a pointer in the form of a bookmark or a Uniform Resource Locator (URL)
    - a bookmark is a pointer that allows the user to "jump" to a specific location in the same document by clicking the anchor
    - a URL is a pointer that enables the user to access other Web resources (i.e. text, image, audio and video files, as well as executable programs and other Internet protocols such as gopher, ftp, telnet and WAIS) in the same or a different server (Figure 4)

1.4.5. Web browsers

- a Web browser is the interface between the user and the Web (Figure 5)
  - in the client/server model of computing of the Web, the browser is a client application
  - the browser allows the client computer to request the service of one or more server computers by means of URLs
  - when a URL is entered in the location field of a browser, the browser goes through the following steps to establish the connection with the server:
    - determine what protocol to use
    - look up and contact the server at the address specified
    - request the specific document from the server
  - when the server has responded, the browser interprets and executes the HTML commands to display the returned text and images on a specific graphical user interface (GUI) platform (i.e. Windows, Macintosh or UNIX)

1.4.6. Plug-ins and helper programs
plug-ins and helper programs are client-side applications developed to enhance the functionality of Web browsers

• a plug-in is a small application program that is installed to enable browsers to perform a specific function not supported by their generic capabilities
  • example: browsers support only certain bitmap or raster formats, plug-ins are required to display vector graphics
  • plug-ins are product-specific, i.e. a particular plug-in is developed and can be used for a specific software application only
    ■ most plug-ins are free from software vendors
    ■ Netscape Navigator and Internet Explorer now contain basic plug-ins which are installed automatically when the browsers are installed
    ■ when the browsers requires a plug-in that has not yet been installed, it will prompt the user to download and install the plug-in (see 1.4.7)
    ■ plug-ins are loaded automatically when the browser is launched

• a helper program is a large program that is installed on the client computer to perform relatively complex custom applications, usually in a separate window (Figure 6)
  • example: browsers do not support video files, helper programs are required to display video files
  • helper programs are product-specific, i.e. a particular helper program is developed and can be used for a specific software application only
    ■ helper programs are usually free from software vendors
    ■ helper programs are installed separately and are loaded automatically when they are needed by the browser (see 1.4.7)

1.4.7. Multiple Internet Mail Extension

• when a Web browser receives a file, it is able to determine whether the file is readable by itself or the file needs to be passed on to a helper program or plug-in
• this is achieved by the method of Multiple Internet mail Extension (MIME)
  • MIME defines the type of files used on the Web
    • all HTML, images, video and audio files have a specific file type and specific file extension
      ■ for example, html and htm for HTML files; gif and jpg for graphics files; au, wav and mp2 for audio files and mpeg, mov and avi for video files)
    • the server uses the MIME information in its configuration to figure out the types of the files it sends to the browser
    • the browser in turn uses the MIME information in its configuration to determine what it needs to do in order to display the files it receives from the server
• in this way, no matter where the browser gets a file, it can always figure out how to display its contents either using its own functions or with the aid of a helper program or a plug-in

2. Surfing the World Wide Web for Information
in order to serve information in the Web, it is necessary to
- set up the server computers properly
- organize the Web documents
- install a Web browser on the client computer

2.1 The Three-tiered Client/server Architecture of the Web

- in its simplest form, the client/server model of computing is a two-tiered architecture that is made up of a client computer and a server computer
- in the Web, vendors are now using gateways and Common Gateway Interface (CGI) to connect Web browsers to virtually all forms of client/server systems, including databases, transaction processing, e-mail backbones and graphics
  - a gateway is a special piece of software in a server computer that allows it to run application programs called CGI scripts
  - these application programs are usually referred to as "back-end" programs and CGI is the protocol in which back-end program are written
  - a back-end program is a server-side application
    - when a server receives a request it does not understand (e.g. data in an input form), it invokes the program to take care of the request
    - the back-end program executes the request and returns the result in HTML format to the Web server, which then treats the result just like a normal document and returns it to the requesting client
  - thus, the Web is practically built on a three-tiered client/server architecture (Figure 7)
    - the client (Web browser)
    - the HTTP or Web server
    - the application server (can be in the same or different computer where the HTTP server resides)
- Web-based geographic data processing applications are built on this three-tiered client/server architecture
  - in the GIS domain, the application server may be referred to as the GIS server or map server
  - GIS servers are product-specific, i.e. they are developed only for particular GIS software products

2.2. People who make a Web site work

- three groups of people are responsible for making a Web site work
  - Web page authors/designers create Web documents using HTML editors and other design tools
  - Webmasters are responsible for the information base of a particular Web site by
    - creating and maintaining internal standards and protocols
    - organizing Web documents (e.g. in different folders/directories according to functions)
  - systems administrators are responsible for configuring, installing and maintaining an HTTP server by
    - identifying a server most suitable for the purpose of an organization
registering the server so that it can be found by other users of the Internet

2.3. Organization of documents in the World Wide Web

2.3.1. Organization of documents in a Web site

- a Web site is a server computer on the Internet that contains one or more Web presentations or documents
- a Web presentation is a collection of one or more Web pages linked together in a meaningful manner, which as a whole describes a body of information or creates some specific effects in the browser of client computers (Figure 8)
- a Web page is a single element of a Web presentation
  - a Web page that serves as the entry or starting point for a Web presentation is called a home page
  - Web pages in a presentation can be linked in different ways using hyperlinks
    - hierarchical or menu structure (Figure 9)
    - linear or sequential structure (Figure 10)
    - Web or free-flowing structure (Figure 11)

2.3.2. Organization of documents in the Web

- documents posted in the Web are logically organized as directories or indexes
- a Web directory is a list of links organized into categories and sub-categories that enable users to find the sites of interest easily and quickly
  - the World Wide Web Virtual Library is the first Web directory developed (Figure 12)
    - it organizes links to Web sites in a library catalog style, starting with categories such as agriculture, computer science, communications and media, education, engineering, etc., and then breaking them down into sub-categories until individual sites are listed
    - it is maintained as a cooperative effort by volunteers all over the world
  - Yahoo and Excite are two examples of commercially maintained Web directories
  - there are hundreds of smaller directories that list Web sites based on special interest (e.g. business, education, entertainment)
- a Web index is a collection of all the Web presentations on the Web
  - the index itself is a database of Web presentations that exist on the Web
    - these Web presentations are collected by using a Web robot (also known as a crawler, worm or spider) that wanders the Internet on its own, jumping from link to link and collecting Web pages for the database
    - the best robots can get an updated collection of the entire Web in about a week's time
    - a particular Web index is used in conjunction with a search engine which is capable of finding and linking specific Web pages by key words (see 2.4.1)
      - examples of Web indexes are AltaVista and Lycos
      - there are many other Web indexes maintained by commercial and non-commercial organizations

2.3.3. Getting included in Web directories and indexes
people who want to have their Web presentations included in a Web directory or index can do so by submitting the URLs of the presentations to the directory or index
- the methods of submission are different for different directories and indexes
- detail of adding a URL can be found in the Web sites of the organizations that maintain the directories and indexes
- alternatively, they can use Web announcement services that will register their Web presentations with all the major Web directories and indexes
  - Submit It! is an all-in-one submission service that will register a Web site with up to 400 search engines and directories

2.4. Search Engines and Search Agents

- in order to find information in the Internet, it is necessary to use a special program called a search engine or a search agent
- conventionally search engines have been designed for searching text-based information but there are now search engines developed specially for searching image- or graphics-based information

2.4.1. Search engines

- a search engine is an application program that operates within a Web browser (Figure 13)
- its function is to enable the user to find information about a specific field of knowledge on the Internet
  - for directory-based search engines, the user starts by selecting a category and then following the links to the specific sub-category of interest
  - the following search engines are based on the use of directories
    - Excite
    - Looksmart
    - Yahoo
  - for index-based search engines, the user starts by entering one or more key words
  - the search engine returns a list of Web pages containing words that match the key words, together with their respective URLs
  - the following search engines are based on the use of indexes
    - AltaVista
    - Hotbot
    - InfoSeek
    - Lycos
    - WebCrawler
- different search engines are different in terms of user interface, search options and resources that they can access
  - as a result, different search engines will return a different number of "hits" or "matches" for the same search key word
- search engines have been used mainly to find text-based information, but it is now possible to find graphics-based information about the location of a business, a city or a country
  - all of the search engines listed above have a "map", "road map" or "travel"
function that is capable of generating a location map according to an input address
- it is also possible to obtain geographic information, in the form of maps and textual descriptions, by using place names and the word "map" as the key words in a search

2.4.2. Search agents

- search agents are also called *meta-search engines* and *searchbots*
- some search agents work within the Web browser but others are stand-alone applications that work outside the Web browser
- when the user has entered the search key words
  - the search agent will make use of a multitude of search engines to search information pertaining to the key word
  - it will then list the top matches from the returns of the search engines
  - the user can view the document in a Web browser by clicking the appropriate items in the returned list of URLs
- an example of search agents is *Copernic* (**Figure 14**), which can be downloaded from the following home page at [http://www.copernic.com/](http://www.copernic.com/)

2.4.3. Image search engines

- an image search engine is a special type of search engine designed for searching image databases on the Internet
  - one example is Webseek developed at Columbia University (**Figure 15**)
    - this is a content-based multimedia search engine that allows the user to search images as well as audio and video files
  - the search engine *Lycos* noted above has a special function that enables the user to search image database in the Internet by specifying a key word
    - this is a multimedia search function (i.e. it is capable of searching image as well as audio and video files)
    - this image search function is activated by clicking the [Pictures and Sound] button in the Lycos interface

2.4.4. Geographic information search engines

- these are search engines specially developed to find geographic information on the Internet
  - instead of using key words, these search engines use geography or location as the search criteria (**Figure 16**)
  - spatial search criteria may include: city names, street addresses or clickable image maps
- examples of geographic information search engines:
  - BigBook, a search engine that makes use of an address and related spatial search criteria to find locations of businesses
  - *Wilkins Tourist Maps of Australia*, a collection of maps and descriptive information about Australia
  - CityGuide, a search engine to access maps and descriptive information about
2.4.5. Map generators and real-time map browsers

- map generators are high-end geographic information search engines that allow the user to find a location and generate a map on the fly
- the user enters specifications such as location, thematic layers and symbology by clicking an image map or using a data entry form (Figure 17)
  - a gateway at the server computer passes the requests to a map or GIS server where the map is composed
  - the resulting map is sent back to the client computer where it can be viewed using native browser capabilities
- examples of map generator:
  - TIGER Mapping Services, US Bureau of the Census
  - Master Environmental Query, US Department of Defense
  - Community Atlas of Canada
- some advanced map generators allow the user to interact with remote geographic databases, these are sometimes labeled as real-time map browsers to distinguish them from the less capable regular map generators
- examples of real-time map browsers:
  - VISA Corporation ATM Locator
  - The Australian White Pages
  - Map Room, City of Oakland, CA

3. Summary

- the World Wide Web is now the de facto standard interface to the Internet
- huge amount of information covering practically all fields of humanities, sciences and technologies can now be found in the Internet
- despite advances in search engine technology, searching for information in the Web is not always an easy task

4. Review and Study Questions

1. Explain the following terms in the context of the World Wide Web:
   - client/server computing
   - Hypertext Markup Language (HTML)
   - plug-in and helper programs
   - Uniform Resource Locator (URL)
   - Web page and home page

2. Define "client pull" and "server push" as applied to the distribution of information across the Internet? List some applications to which these techniques can be used. What are their
relative merits and limitations?

3. Describe the procedures for posting a research paper on the Web.

4. Using the search engines noted in this unit, find information pertaining to the term "geographic information science".
   - Do all search engines return the same number of matches?
   - Are the matches returned to you in an orderly manner?
   - Which search engine in your opinion is the most informative? Justify your choice.

5. Using any of the search engines noted in this unit, find:
   - the campus map of the University of California, Santa Barbara
   - a map showing the Chinatown of San Francisco, CA
   - an aerial photograph or image of Victoria Harbor, Hong Kong

---

**Citation**

To reference this material use the appropriate variation of the following format:


Unit 148 - World Wide Web Basics

Table of Contents

Advanced organizer

Learning outcomes

1. What is the World Wide Web?
   1.1 The World Wide Web and the Internet
   1.2 Who runs the World Wide Web?
   1.3 What is available on the Web?
   1.4 Concepts and definitions of terminology pertaining to the Web
      1.4.1 Client/server computing
      1.4.2 Hypertext Transmission Protocol
      1.4.3 Hypertext Markup Language
      1.4.4 Hypertext and hypermedia
      1.4.5 Web browsers
      1.4.6 Plug-ins and helper programs
      1.4.7 Multiple Internet Mail Extension

2. Surfing the World Wide Web for information
   2.1 The three-tiered client/server architecture of the Web
   2.2 People who make a Web site work
   2.3 Organization of documents in the World Wide Web
      2.3.1 Organization of documents in a Web site
      2.3.2 Organization of documents in the Web
      2.3.3 Getting included in Web directories and indexes
   2.4 Search engines and search agents
      2.4.1 Search engines
      2.4.2 Search agents
      2.4.3 Image search engines
      2.4.4 Geographic information search engines
      2.4.5 Map generators and real-time map browsers

3. Summary

4. Review and study questions

Citation
Unit 148 - World Wide Web Basics

Metadata and Revision History

Page contents

1. About the main contributor
2. Details about the file
3. Key words
4. Index words
5. Prerequisite units
6. Subsequent units
7. Revision history

1. About the main contributor
   
   • author: Albert K. Yeung, Ontario Ministry of Northern Development and Mines, Canada

2. Details about the file
   
   • unit title: World Wide Web Basics
   • unit key number: Unit 148

3. Key words

4. Index words

5. Prerequisite units

6. Subsequent units

7. Revision history
- First draft completed: January 15, 1999
- Posted to website: August 6, 2000

Back to the Unit
**Figure 1: The Concept of Client/Server Computing**

A client can request information from several servers at the time.

A server can serve several clients at the time.

A K. Yeung 1999-02-05 ul48-01
Figure 2: A Web Document in HTML

```html
<HTML>
<HEAD>
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=windows-1252">
<META NAME="Generator" CONTENT="Microsoft Word 97">
<META NAME="Template" CONTENT="C:\PROGRAM FILES\MICROSOFT OFFICE\OFFICE\html">
</HEAD>
<BODY LINK="#0000ff" VLINK="#800080" BGCOLOR="#99ccff">
<B><FONT FACE="Comic Sans MS" SIZE=5 COLOR="#0000ff"><P>Home Page</P>
</B></FONT><P><HR></P>
</B>&nbsp;</P>&nbsp;</P>&nbsp;</P>&nbsp;</P>
</BODY>
</HTML>
```

A.K. Yeung 1999-02-05 u148-02
Figure 3: The Concept of Using Hyperlinks

A.K. Yeung 1999-02-05 u148-03
Figure 4: Components of a Universal Resource Locator (URL)

http://www.anysite.net/data/maps/index.html

- Internet protocol
- Domain name of server computer
- Path name on the server's hard drive in which Web document is stored
- Name of Web document

A.K. Yeung 1999-02-05 u148-04
Figure 5: The Web Browser as an Interface to Internet Resources

Accepts URL → Contacts and connects server → Accepts and parses response from server → Displays response from server

A.K. Yeung 1999-02-05 u148-05
**Figure 6: Running a Helper Program**

Helper programs are run in windows outside the Web browser.

A.K. Yeung 1999-02-05 u148-06
Figure 7: The Three-tiered Architecture of the World Wide Web

Client computer

Web browser

Requests

Internet

Results (Data, reports, maps)

Server computer

HTTP (Web) server

Application

specific requests

Application server

Results (Data, reports, maps)

Application programs

HTTP documents

Database

A K. Yeung 1999-02-05 u148-07
Figure 8: Organization of a Web Presentation

A.K. Yeung 1999-0240 u148-08
Figure 9: Hierarchical or Menu Structure of Web Presentation

A.K. Yeung 1999-02-05 u148-09
Figure 10: Linear or Sequential Structure of Web Presentation

Home page

A.K. Yeung 1999-02-05 u148-10
Figure 11: Web Structure of Web Presentation

A.K. Yeung 1999-02-05 u148-11
Figure 12: The World Wide Web Virtual Library
**Figure 13: A Search Engine Working within a Web Browser**

A.K. Yeung 1999-02-05 u148-13
Figure 14: Interface of the Copernic Search Agent
Figure 15: Interface of Webseek --- an Image Search Engine

A.K. Young 1999-02-05 u148-15
Figure 16: Interface of CityGuide --- an Geographic Search Engine
Figure 17: Example of a Map Generator - TIGER Mapping Services