

World-Wide Web = WWW

Introduction

The World-Wide Web: summary

The following gives an overview of the World-Wide Web:

1. **Introduction:**
Description / definition of the WWW
and its relation with the Internet
2. **Client programs that allow you to use the WWW**
3. **How to save selected information from the WWW to
your computer?**
4. **The success of the WWW**

The World-Wide Web: prerequisites

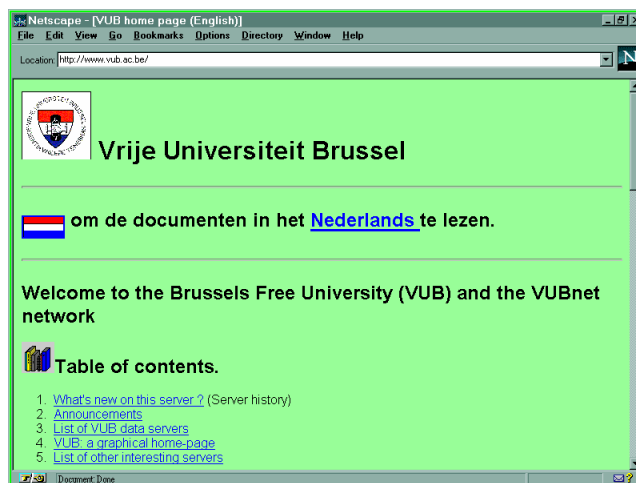
Before using the WWW you should ideally already have learned to understand and to use

- computer hardware
- computer software
- the Internet
- older methods for online communication, such as telnet



****Example

The WWW: example of a welcome page



?? Question ??

Indicate some difference between telnet and the World-Wide Web?



The WWW: views on information space

- Unlike telnet and ftp, the WWW offers a view on one virtually unified but decentralized information space.
- Selecting a server, and switching from one server to another is easier than with telnet and ftp.
- A client program for the WWW can be used to search for information held on a distributed network of WWW hosts / servers.
- A WWW client has a seamless view of the information, even though this information is distributed over many different hosts.

URL = Universal Resource Locator

- = draft standard for specifying an object on the Internet
- the structure is in most cases
protocol://computer_address[/path_name/file_name]
- examples:
 - » <telnet://biblio.vub.ac.be>
 - » <ftp://ftp.vub.ac.be/>
 - » <gopher://gopher.vub.ac.be/>
 - » <http://www.vub.ac.be/BIBLIO/index.html>
 - » <news://news.server.edu/comp.infosystems.www>

URL format / structure

1. The first part of a URL, before the colon “:”, specifies the access method = protocol
2. The second part of the URL, after the colon “:”, is interpreted specific to the access method.
In general, two slashes after the colon indicate a machine /computer name.

?? Question ??

A Uniform Resource Locator or 'URL' is (select 1):

- A browser feature that helps you find web pages
- A term for managing and locating web addresses
- A technical term for a web address
- The American Army and Navy website



?? Question ??

In a URL, what would the part "index.shtml"?

(Select one)

- A directory?
- A file?
- A domain name?



WWW = World Wide Web = W3: description

- **The WWW provides a hypertext-hypermedia interface to information resources in the Internet.
So WWW is a distributed hypermedia system.**
- **The basis of the WWW was developed at CERN (the European Laboratory for Particle Physics) in Switzerland by Tim Berners-Lee and co-workers.**

?? Question ??

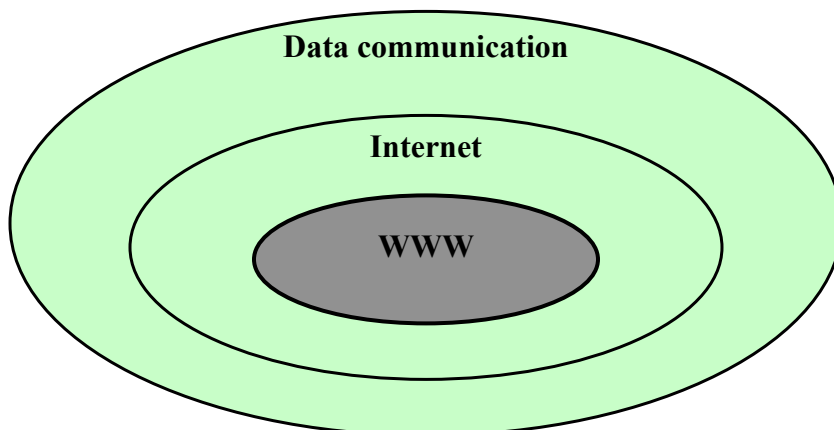
**What is the difference between
Internet and the World-Wide Web?**




The WWW is an application of Internet

- The World-Wide Web (WWW) is a service, an application of Internet.
- It is based on the Internet infrastructure.
- So the WWW is newer than the Internet.
The concept of the WWW was created at the end of the 1980s when the Internet was already well established.

The WWW is an application of Internet: scheme



The WWW: the essential elements

- **Information delivery and access using hypertext/hypermedia documents/objects**
 - » html documents
 - » http protocol: http clients  http servers
- **Integration of protocols in the Internet:**
 - » http servers offering html documents including links to other http servers, telnet servers, ftp servers, nntp servers, gopher servers...

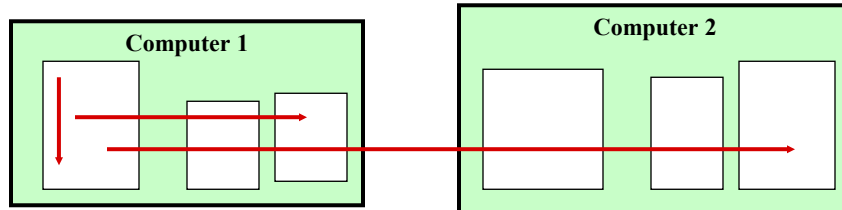
The WWW: function

- **The WWW works by establishing hypertext/hypermedia links between documents anywhere on the network.**
- **A document might include many links to other documents held on many different servers.**
- **Selecting any one of those links will take you to the related document wherever it is.**
e.g. the references at the end of a paper might have hypertext links to the actual documents held elsewhere.

The WWW: hyperlinks

Hyperlinks can link a part of a hypermedia document to

- another part of the same document file
- another document file on the same server computer
- another document file on a server computer located elsewhere in the world



The WWW: hypertext mark-up language = HTML

- Hypertext mark-up language = HTML = the system of codes used by authors to build the hypertext-pages/files in WWW, for instance to create a title or an anchor.
- The codes are invisible / transparent for the user / reader.

!! Task - Assignment !!

Read

HTML. Introductory paper. [online]

Available from:

<http://www.coe.missouri.edu/~DL/iDLR/viewpaper.php?pid=14>

[cited 2005]



?? Question ??

What does HTML stand for? (select one)

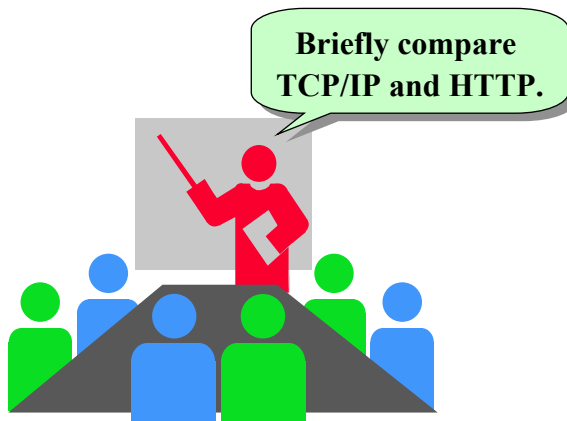
1. Hot Terminal Management Location
2. Hyper Text Machine Language
3. Hyper Text Markup Language
4. Hyper Tradition Men's Lingerie
5. Hot Metal



The WWW: hypertext transfer protocol = HTTP

- Hypertext transfer protocol = HTTP = the software conventions used by client and server programs for WWW to request and transfer hypermedia documents.
- The protocol must not be known by the user / reader = the protocol is invisible / transparent for the user.
- Analogous with the telnet, ftp and gopher protocol.

?? Question ??



The Internet plus WWW or TCP/IP plus HTTP

WWW = *HTTP* +...

Internet = **TCP/IP** +...

The WWW: pages and forms

- **Pages**

Many documents developed for WWW are kept small and are named “pages”.

These often refer to several other “pages”.

- **Forms = gateways to services and databases on server computers in WWW**

Some pages contain electronic forms, to be filled in by the user.

The WWW applications

Analogous to gopher applications:

- Access to online public access catalogues
- Campus-wide information systems
- Access to subject-oriented information
- Access to computer file archives
- Traveling / navigating through the Internet via linked html-pages
- Access to intranets within institutes / companies

The WWW: Which information is offered via the WWW?

- Many types of information formats
- Many types of information contents (= on many subjects)
Examples:

» sources on communications & telecommunications:

<http://www.analysys.co.uk/commslib.htm/>

» the home page of a broadcasting company:

<http://www.vpro.nl/>

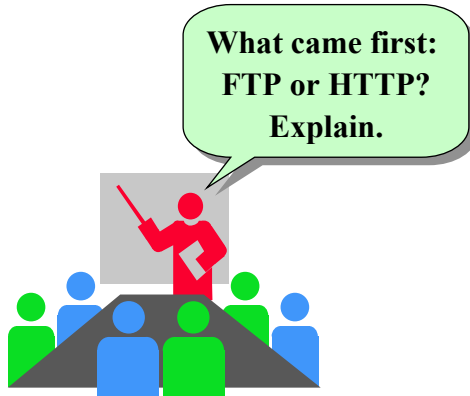
PURL = Persistent URL: introduction

- **Some URLs have the word "PURL" located in the early part of the URL.**
- **To obtain a PURL the owner has had to register the site with an intermediary PURL service. If for any reason the site moves addresses, the owner can register the change of address with the PURL service which then redirects any users to the new URL.**

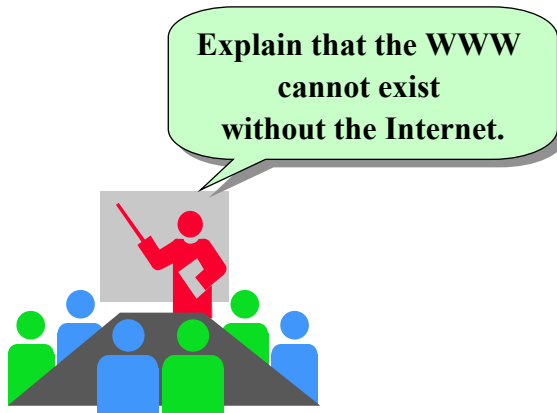
PURL = Persistent URL: advantages

- **A PURL address should not lead to a dead link and should mean that the same URL will always point to the same resource, even if (behind the scenes) the resource has been moved from server to server.**
- **So a PURL is a clue that the owner of the resource is committed to keeping the site stable and persistently available via a given URL.**

?? Question ??



?? Question ??



World-Wide Web = WWW

WWW client programs

WWW: client / browse programs

- **To access the WWW, you run a browser program.**
- **The browser reads documents, and can fetch documents from other sources. Information providers set up hypermedia servers which browsers can get documents from.**
- **The browser can display hypertext documents. Hypertext is text with pointers to other text. The browsers let you deal with the pointers in a transparent way: select the pointer, and you are presented with the text that is pointed to.**

WWW: types of client programs (browsers)

To access WWW, you can use

- » a browser on a network server computer that can be telnetted to, or
- » a browser on your own machine
(for instance the microcomputer you use)

WWW: browsers accessible by telnet

- Example of client programs (browsers) to access WWW, running on Unix computers, in character mode: *lynx*
- Example of a public access browser accessible by telnet: at the following server computer: `info.cern.ch`
No password is required.

WWW client program on Unix, in character mode: lynx

Lynx default home page (p1 of 2)

WELCOME TO LYNX AND THE WORLD OF THE WEB

You are using a WWW Product called Lynx. For more information about obtaining and installing Lynx please choose About Lynx

The current version of Lynx is 2.3. If you are running an earlier version PLEASE UPGRADE!

INFORMATION SOURCES ABOUT AND FOR WWW

- * For a description of WWW choose Web Overview
- * About the WWW Information Sharing project
- * WWW Information By Subject
- * WWW Information By Type

OTHER INFO SOURCES

-- press space for next page --

Arrow keys: Up and Down to move. Right to follow a link; Left to go back.

H)elp O)ptions P)rint G)o M)ain screen Q)uit /=search [delete]=history list

WWW: browsers for your own computer

- The preferred method of access of the Web is to run a browser yourself on your computer.
- These provide not only an interface to http-servers in WWW, but also to various other protocols

- » ftp
- » gopher
- » SMTP, POP, IMAP (for e-mail)
- » NNTP (for Usenet News)
- » ...



WWW: examples of browsers for your own computer

**Browsers are available for many computer platforms;
in particular:
browsers for *Windows*, in historical order:**

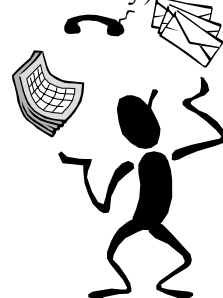
- » *Netscape*
- » *Microsoft Internet Explorer*
- » *Opera*
- » *Mozilla Firefox*
- » ...



Netscape: a WWW browser program

***Netscape* is a client program for**

- » http servers
- » ftp servers (downloading only, 1 file at the time only)
- » gopher servers
- » SMTP, POP,... (for e-mail)
- » NNTP servers (for Usenet News)
- » ...



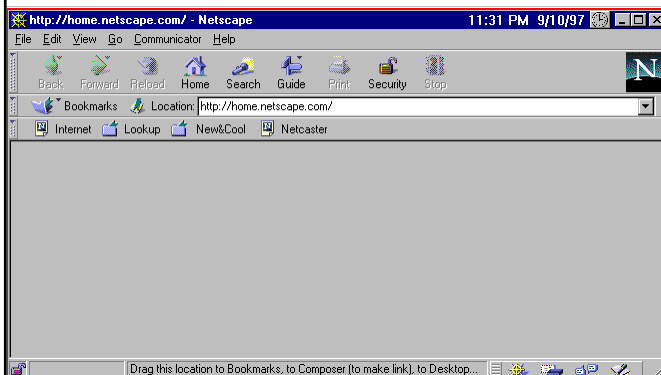
Netscape: program characteristics

Netscape versions exist for several computing platforms:

- » UNIX X-Windows
- » Apple Macintosh
- » DOS + Microsoft Windows + WinSock
- » Windows 95,...
- » Windows NT, 2000,...
- » ...



Netscape for Windows 95: screen shot



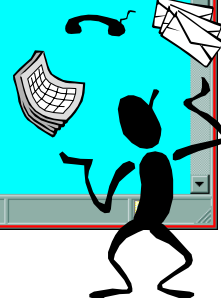
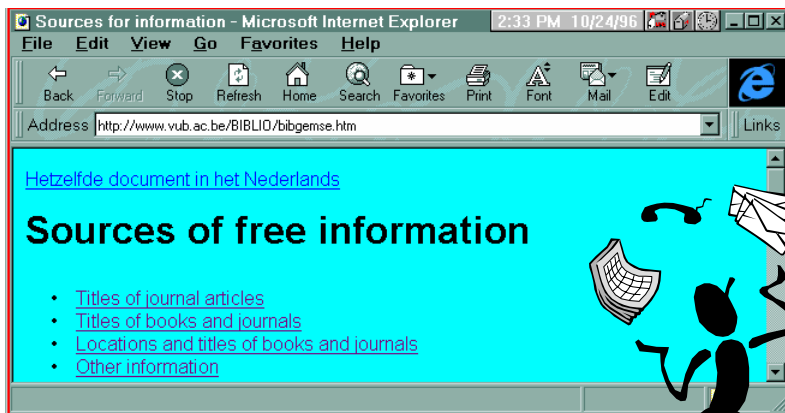
MS Internet Explorer: a WWW browser program

This is a suite of client programs for

- » http servers
- » ftp servers
- » gopher servers
- » SMTP, POP,... (for e-mail)
- » NNTP servers (for Usenet News)
- » ...



MS Internet Explorer 3 for Windows 95: screen shot



MS Internet Explorer 4 for Windows 95: screen shot



!! Task - Assignment !!

Learn the basics
of (hyper)links and (WWW) browsers through
the WebWise online course that is available from
<http://www.bbc.co.uk/webwise/course/>



!! Task - Assignment !!

Learn the basics of WWW, by reading
Guide to Network Resource Tools (GNRT). [online]
Available from: <http://gnrt.terena.nl/> [cited 2004]
WWW



!! Task - Assignment !!

Learn the basics of the WWW, by reading
Cohen, Laura
Understanding the World Wide Web. [online]
Available from:
<http://www.internettutorials.net/www.html>
[cited 2006]



!! Task - Assignment !!

Read:

Brain, Marshall

How web servers work. [online]

Available from:

<http://computer.howstuffworks.com/web-server.htm>

[cited 2005]



?? Question ??

Which differences do you know
between
the major WWW client programs?



Differences between the WWW clients by *Netscape* and *Microsoft* (Part 1)

Differences among the competing programs with the same version number:

- » Direct copy and paste through *Windows* of images works with *Microsoft IE 4*, but NOT with *Netscape Navigator 4!*
- » ActiveX controls are mainly used with *Microsoft* software and do not work with *Netscape*, unless a specific plug-in is installed!

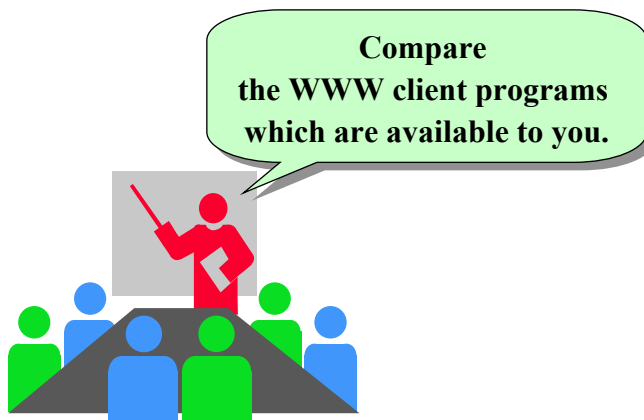
Differences between the WWW clients by *Netscape* and *Microsoft* (Part 2)

- » *Windows* software installed on the pc can be updated easily online using *Microsoft IE 5*.
- » It is easier to copy and use previously downloaded files from the cache on disk with *Microsoft IE 4* than with *Netscape Navigator 4* .

Differences between the WWW clients by *Netscape* and *Microsoft* in printing

- » Printouts are often better with *Netscape Navigator 4* than with *Microsoft IE*, at least up to version 5!
 - nicer borders and characters
 - tables, paragraphs and lines of text not split at the end of the page
- » *Netscape Navigator 1 to 4* gets the document again from the Internet before printing it, so that it prints slower than *Microsoft IE 4* in most cases.

!! Task - Assignment - Exercise !!



?? Question ??

**Which client program
do YOU use or will YOU use
to access the WWW?**



!! Task - Assignment - Exercise !!

**Browse the WWW,
using an available
browser client program.**



?? Question ??

Select the correct one:

1. Hyperlinks look always like underlined text
2. A hyperlink can be in a text or in an image
3. Hyperlinks are always coloured blue



?? Question ??

How can you print a web page? (select one)

1. First you must load them in your word processing program
2. You can use the print functions in the browser software
3. You can print from the browser, but only if you have a colour printer
4. You cannot print web pages



!! Task - Assignment - Exercise !!

**Visualise the HTML source code
of a WWW page,
using a WWW client program.
What do you learn from this exercise
about the basic properties of HTML?**



!! Task - Assignment - Exercise !!

**Exploit the possibility
to open more than one window,
using a WWW client program
in Windows.**



?? Question ??

Why would you want to open more than one window on WWW servers, using a WWW client program?



!! Task - Assignment !!

What means “caching” in the context of using a web. Where else is caching used?



!! Task - Assignment - Exercise !!

**Decide on the size of the web cache
on your personal computer.**

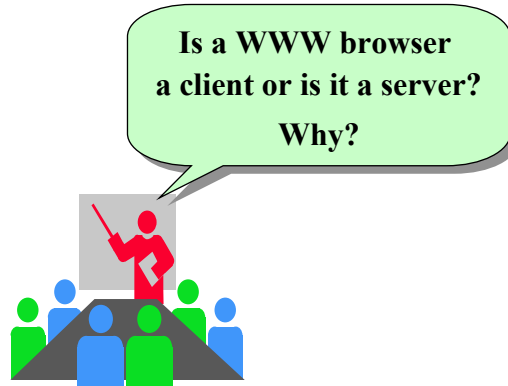


?? Question ??

**What is HTTP and HTML?
Explain the difference and the relation
between both.**



?? Question ??



Programs to access the Web and other Internet services

- The WWW has become a tremendous success in the 1990s.
- It has made the Internet popular.
- The programs to access the WWW have incorporated many functions to exploit other services offered by the Internet.
- A consequence is that for many users the distinction between Internet and WWW is blurred.

?? Question ??

**What came first: Internet or WWW?
Explain.**



World-Wide Web = WWW

Saving information from a web

WWW: How to save information from a web?

Information displayed by your web browser/client program can be saved,

- by select, copy, paste in another document (and save)
- by saving a complete page to your disk
 - » in separate files
(for instance 1 HTML file + some image files)
 - » in 1 file, using *Microsoft Internet Explorer 5* or a later version
- by copying the information into an e-mail message that you send to your own e-mail account

!! Task - Assignment - Exercise !!

Copy some text fragment from WWW and paste it into another document on your computer.



!! Task - Assignment - Exercise !!

Save a text from WWW
to disk, as HTML,
using a browser program.



!! Task - Assignment - Exercise !!

Display an HTML file
that you have saved
from the WWW to your disk,
in a program for word processing.
Is the file displayed properly?



!! Task - Assignment - Exercise !!

Can the Internet/WWW program that you use
copy a picture from WWW,
so that you can directly
paste it into a document
in another program on your PC.



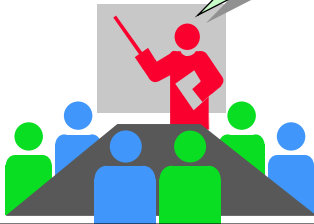
!! Task - Assignment - Exercise !!

Save a picture from WWW
to disk,
using a browser program.



!! Task - Assignment - Exercise !!

Check if the program that you use
for word processing
allows you to insert a picture
that you saved to disk
into your word processing document.



WWW: How to save a HTML document including pictures?

Saving a complete HTML document including pictures can be done by using the appropriate software.

For instance:

- » *Netscape Page Composer*
(included in the *Netscape* software suite)
- » *(Microsoft Internet Explorer 4 with FrontPage Express)*
- » *Microsoft Internet Explorer 5, or a more recent version !*

!! Task - Assignment - Exercise !!

**Save a document that includes at least 1 image
from WWW to your disk,
and test if the images are saved too,
by loading the saved HTML file
in a program that can show HTML files.**



!! Task - Assignment - Exercise !!

**Choose a web page with images.
Save this as HTML.
Look at the files on your disk, that were saved there.
How many and which kind of files do you observe?**



!! Task - Assignment - Exercise !!

Why is saving all kinds of web pages more complicated than saving for instance a page with graphics that has been composed with a program for word processing?



!! Task - Assignment - Exercise !!

**Choose a web page with images.
Save this as 1 archive file,
using *Microsoft Internet Explorer 5*
or a more recent version.
Open this saved archive file
and see if the result corresponds well
to the original document.**



!! Task - Assignment - Exercise !!

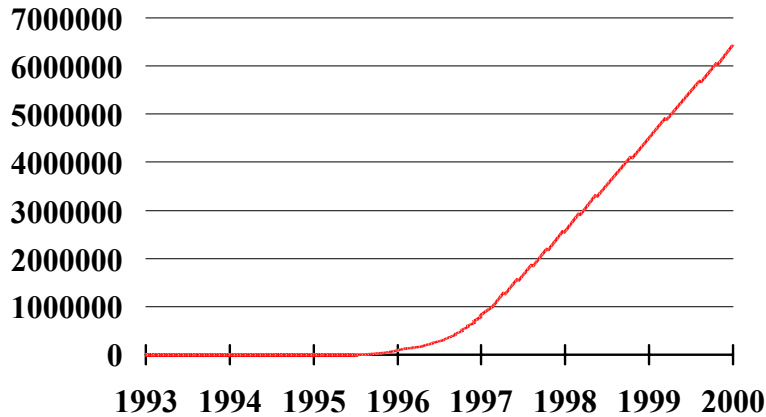
**Download / save a page and also all linked pages
from a web site,
so that these are later directly available to you,
even off-line.**



World-Wide Web = WWW

The success of WWW

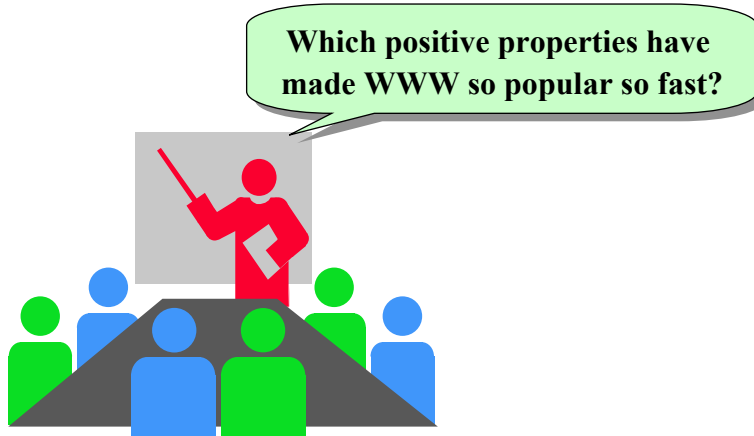
WWW: growing number of WWW servers



WWW as popular method to access information from computers

- The WWW has quickly become the most popular medium to access information that resides on various computers that are connected to a computer network.

?? Question ??



WWW properties leading to its success, related to access (Part 1)

- The client-server architecture that is used in the web allows the user to choose a browser client program.
- Browser client programs are available free of charge.
- The WWW is accessible from computers with most operating systems.
- The WWW offers an easy to use interface to various complex services, based on hypertext/hypermedia.
- The multimedia contents are attractive.

WWW properties leading to its success, related to access (Part 2)

- **Access to multimedia contents is easy, because browser programs include viewer software for several formats.**
- **The WWW makes it easy to switch from one server to another transparently (“browsing / navigating /surfing”).**
- **Subject trees and indexes exist, to guide the user to relevant information sources.**
- **Much of the information that can be retrieved through the WWW is available free of charge.**

WWW properties leading to its success, related to access (Part 3)

- **WWW browser client programs also offer access to anonymous ftp, gopher, WAIS and other databases which were made accessible through WWW, e-mail, Usenet newsgroups,...**

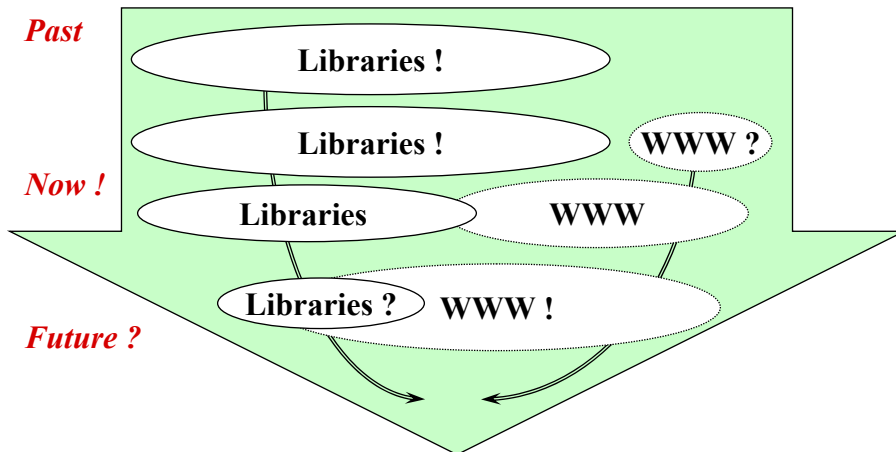
WWW properties leading to its success, related to publication (Part 1)

- **The WWW is rather well standardised, but still evolving and improving.**
- **The WWW is based on an *open*, published and freely applicable standard / protocol (named “http”).**
- **Some good http server software packages are available free of charge.**
- **Some good programs to create information pages suitable for use through the WWW, with HTML, are available free of charge.**

WWW properties leading to its success, related to publication (Part 2)

- **It is relatively easy to create HTML pages.**
- **Not only text, but multimedia contents can be published.**
- **The WWW allows the incorporation of Internet information sources and services that already existed before the WWW; for instance anonymous ftp, gopher, WAIS, electronic mail,...**

WWW and libraries as institutions



- You are free to copy, distribute, display this work under the following conditions:



» Attribution:

You must mention the author.



» Noncommercial:

You may not use this work for commercial purposes.



» No Derivative Works:

You may not change, modify, alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.