

AN ADVANCED GUIDE TO THE INTERNET

I.T. Works seminar of 27 October 1994

World-Wide Web

A hypermedia infospace built on top of the Internet

Hans C. Arents

Materials Information Processing Systems group

Department MTM, K.U.Leuven, W. de Croylaan 2, B-3001 Leuven, Belgium

Tel: +32 (0)16 32 12 32 Fax: +32 (0)16 32 19 91 E-mail: Hans.Arents@mtm.kuleuven.ac.be



World-Wide Web: A hypermedia infospace built on top of the Internet

What is the WWW?

How the WWW works:

- browser/server architecture
- browser side: URL and HTML
- server side: HTTP, MIME and CGI

How to use the WWW:

- living in the global infospace
- setting up a local infospace

Summary and conclusions



What is the WWW?

The original idea

What is hypermedia?

Global distributed hypermedia system

- *Memex* Vannevar Bush (1945)
- *Xanadu* Theodore Nelson (1981)
- *World-Wide Web* Tim Berners-Lee (1989)

World-Wide Web: ÿ the Internet killer application

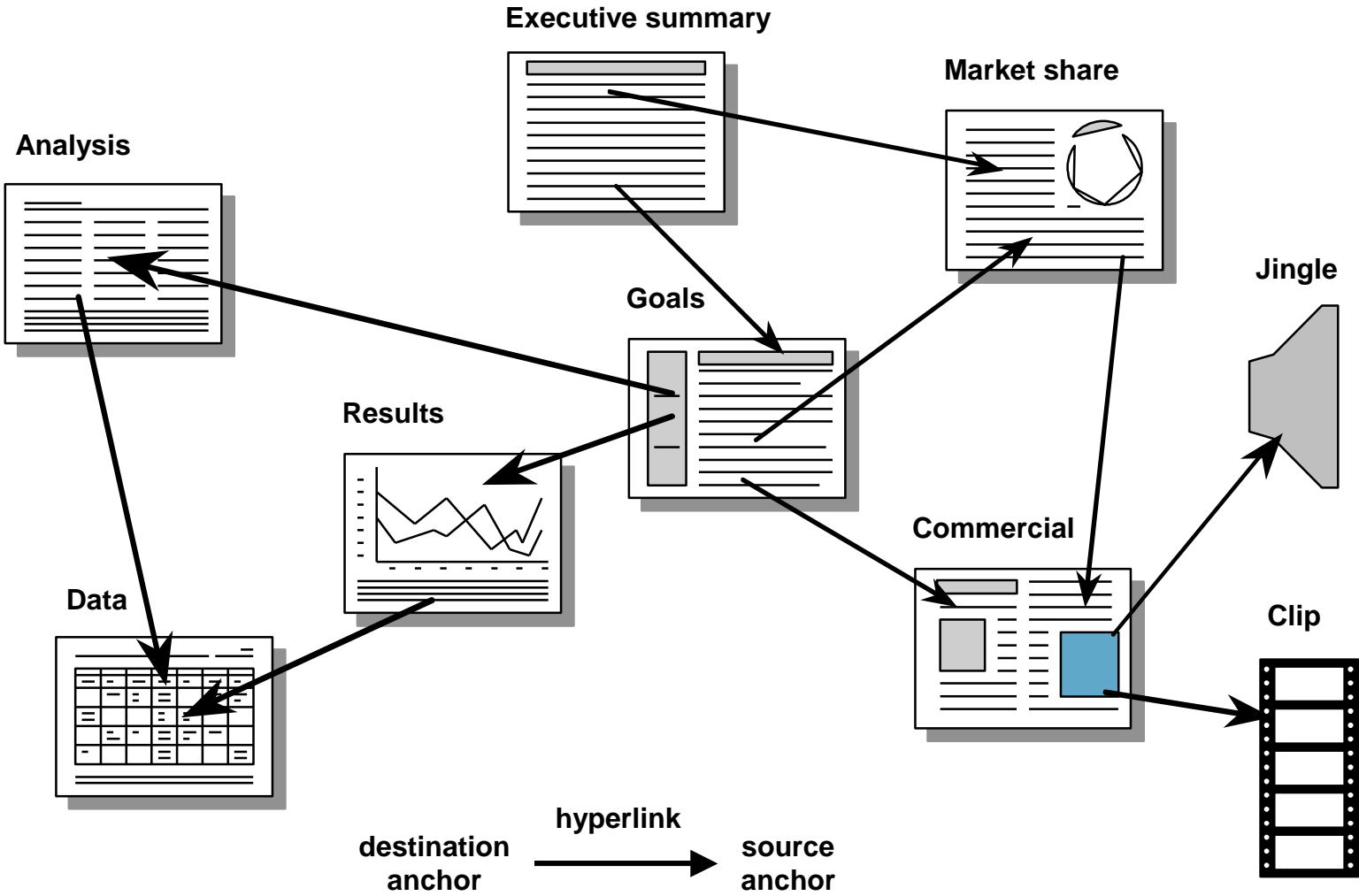
ÿ the *VisiCalc of the '90s*

Internet + World-Wide Web = global hypermedia infospace



What is the WWW?

Hypertext and hypermedia



What is the WWW?

A brief history

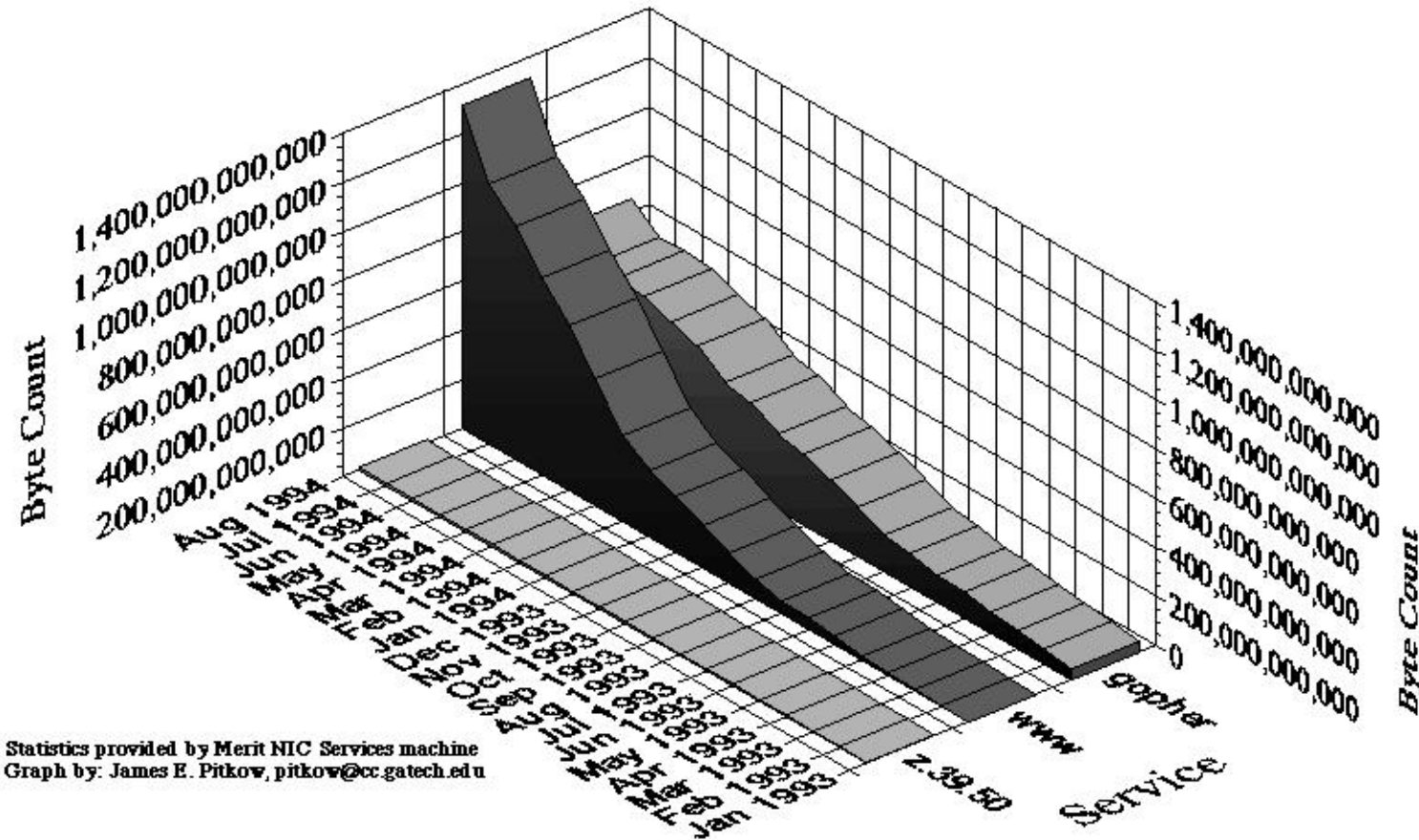
- 1989** Proposal for World-Wide Web (CERN, Tim Berners-Lee)
- 1991** Prototype browser for NeXTStep demonstrated at Hypertext '91
- 1993** *Mosaic* graphical browser for X Windows (NCSA, Marc Andreessen)
- March 1994** WWW byte traffic surpasses Gopher traffic on the NSFnet backbone
- May 1994** 1st International WWW conference at Geneva (sold out in 5 days!)
- October 1994** 2nd International WWW conference at Chicago (over 1300 attendees!)
- November 1994** IBM *OS/2 Warp* with built-in suite of Internet utilities (WebExplorer)
- July 1995 (?)** Microsoft *Windows95* with built-in access to the Microsoft Network



What is the WWW?

Growth of the World-Wide Web

Byte Count by Service



What is the WWW?

Growth of the World-Wide Web

Total number of servers:

- January 1993: **50** - June 1994: **1500** - *today: 5000*

è **10 new servers launched every day**

Total number of browsers: **1.000.000**

è **1000 browsers downloaded to new sites every day**

Total amount of data:

- in 1992: **0.5 Gb** - January to March 1993: **5 Gb** - *today: 10 Gb*

è **3.000 % annual growth or 1 % growth *per day***



What is the WWW?

Users of the World-Wide Web

Top five types of users:

Domain	Percent of WWW traffic	Percent of Internet hosts
U.S. educational (*.edu)	49 %	27 %
U.S. commercial (*.com)	20 %	26 %
U.S. government (*.gov)	9 %	6 %
United Kingdom (*.uk)	7 %	5 %
Canada (*.ca)	5 %	4 %

Typical user:

- between 21 and 30 years old (56 %), male (94 %)
- U.S. citizen (69 %), professional user (45 %) or graduate student (22 %)



What is the WWW?

Usage of the World-Wide Web

Frequency of use:

1 time per week	2%
few times per week	18%
1 to 4 times per day	42%
5 to 8 times per day	18%
9 or more times per day	20%

Since how long on the Web:

a month or less	10%
1 to 6 months	25%
6 months to a year	26%
1 year or more	39%

Hours spent exploring the Web:

under 5 hours per week	42%
6 to 10 hours per week	36%
10 or more hours per week	22%

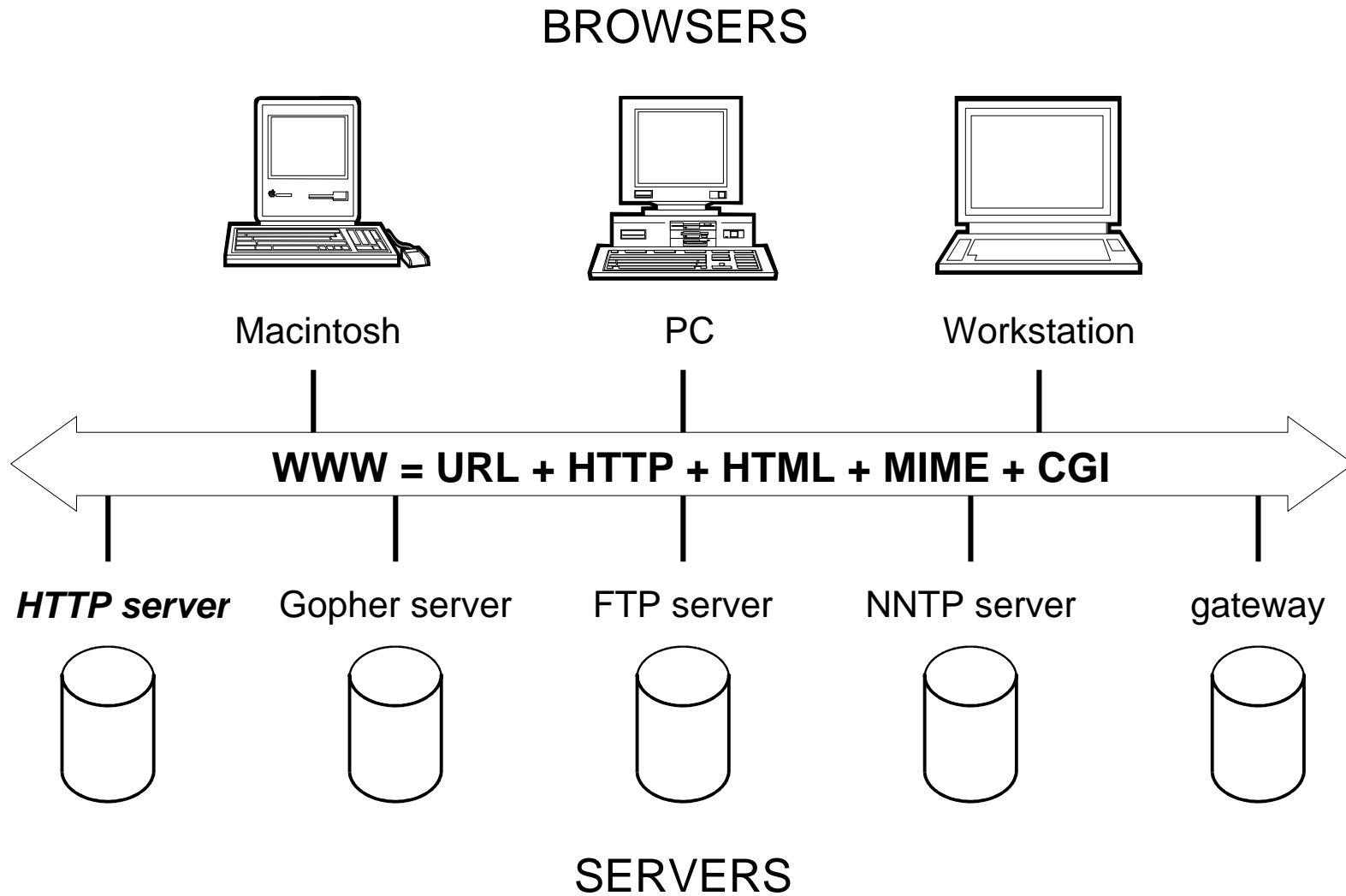
Speed of connection:

less than 14400	3%
14400	37%
faster than 14400	60%



How the WWW works

Browser/server architecture



How the WWW works

Browser/server architecture

Browsers:

- NCSA Mosaic (MS Windows, X Windows, Macintosh - the original browser)
- MCC NetScape (MS Windows, X Windows, Macintosh - most widely used)
- Lynx, WebExplorer, AIR Mosaic, Cello, WinWeb, MacWeb, Netcruiser,...

Servers:

- NCSA httpd (most widely used server)
- CERN httpd (most comprehensive server)
- MCC NetSite (first commercial strength server)

providing access to: WWW, Gopher (Veronica), FTP (Archie), Usenet, WAIS, ...



How the WWW works

Key concepts

CLIENT SIDE:

Addressing scheme	<i>URL</i>	Universal Resource Locators
Content encoding	<i>HTML</i>	HyperText Markup Language

SERVER SIDE:

Transfer protocol	<i>HTTP</i>	HyperText Transfer Protocol
Format negotiation	<i>MIME</i>	Multimedia Internet Mail Extension
Service extension	<i>CGI</i>	Common Gateway Interface



How the WWW works

Addressing scheme

Global search, access and retrieval of documents

- transmission protocols ⌘ network gateways
- storage formats ⌘ format conversion
- *addressing scheme* ⌘ *universal identifiers*

World-Wide Web addresses: extensible, complete, printable

- URI **U**niversal **R**esource **I**dentifier
- URL **U**niversal **R**esource **L**ocator
- URN **U**niversal **R**esource **N**ame



How the WWW works

URI, URL and URN

URI: Universal Resource Identifier

- syntax for encoding arbitrary addressing schemes
- RFC 1630

URL: Universal Resource Locator

- syntax for encoding actual addresses of data
- WWW standard

URN: Universal Resource Name

- syntax for encoding conceptual addresses of data
- under development



How the WWW works

URL: Universal Resource Locator

AccessMethod : AccessPath

AccessMethod : // AccessServer / AccessLocation

Examples:

• http <http://dept.uni.edu:80/directory/file.html>

• ftp <ftp://ftp.dept.uni.edu/directory/file.ps.Z>

• file file://localhost/directory/file

• gopher gopher://dept.uni.edu/

• wais wais://dept.uni.edu/

• mailto <mailto:somebody@dept.uni.edu>

• news news:alt.hypertext



How the WWW works

HTML: HyperText Markup Language

Design

- 8-bit ASCII with limited set of tags inserted in-between
- easy to learn by novices, but extensible, powerful and robust
- an application of SGML (Standard Generalized Markup Language)
- formatting and presentation is exclusive responsibility of the browser

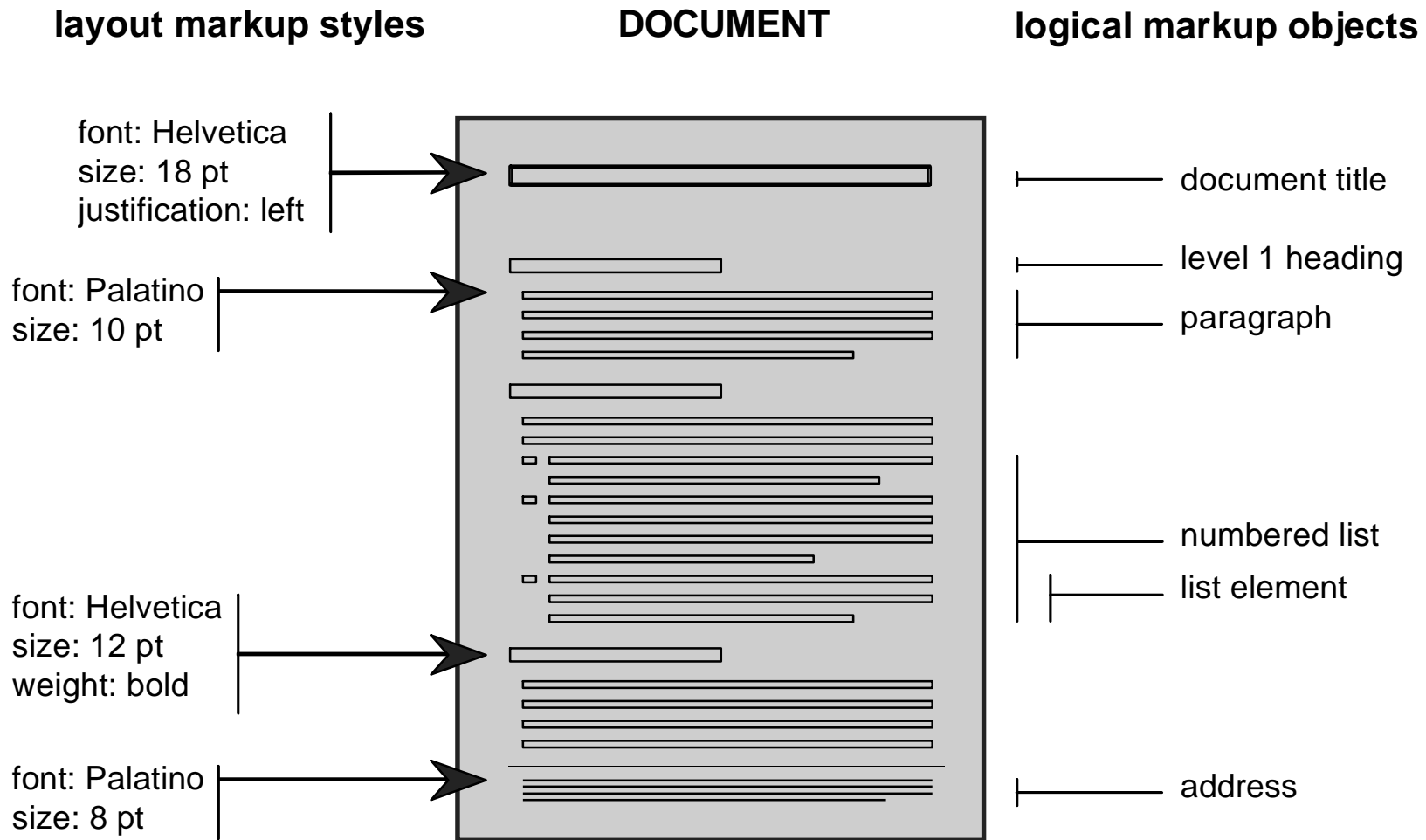
Use

- describe structure of the hypermedia documents
- define hypertext links and anchors
- refer to in-line images



How the WWW works

HTML - Logical markup



How the WWW works

HTML - Example: a document

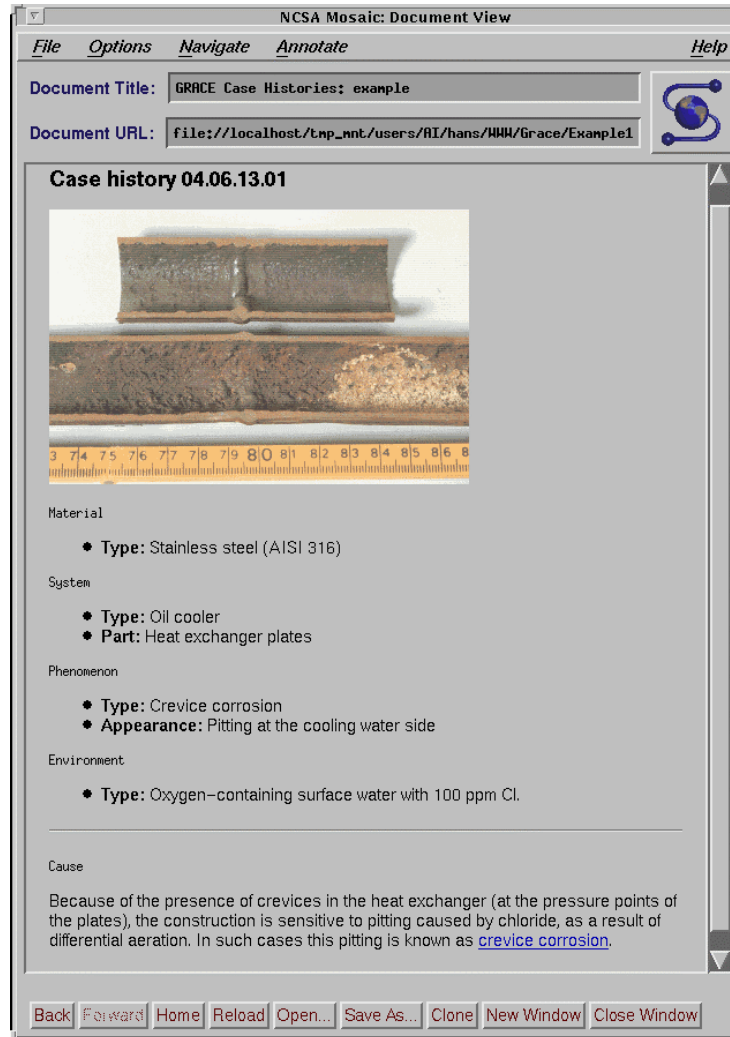
```
<html>
<head>
<title>GRACE Case Histories: example</title>
<link rev="made" href="mailto:hca@mtm.kuleuven.ac.be">
</head>
<body>
<h1>Case history 04.06.13.01</h1>

<h2>Material</h2>
<ul>
<li><b>Type:</b> Stainless steel (AISI 316)</li>
</ul>
<h2>System</h2>
<ul>
<li><b>Type:</b> Oil cooler</li>
<li><b>Part:</b> Heat exchanger plates</li>
</ul>
<h2>Phenomenon</h2>
<ul>
<li><b>Type:</b> Crevice corrosion</li>
<li><b>Appearance:</b> Pitting at the cooling water side</li>
</ul>
<h2>Environment</h2>
<ul>
<li><b>Type:</b> Oxygen-containing surface water with 100 ppm Cl.</li>
</ul><hr>
<h2>Cause</h2>
<p>Because of the presence of crevices in the heat exchanger (at the pressure
points of the plates), the construction is sensitive to pitting caused by
chloride, as a result of differential aeration. In such cases this pitting
is known as <a href="Crevice.html">crevice corrosion</a>.</p>
</body>
</html>
```



How the WWW works

HTML - Example: a document



How the WWW works

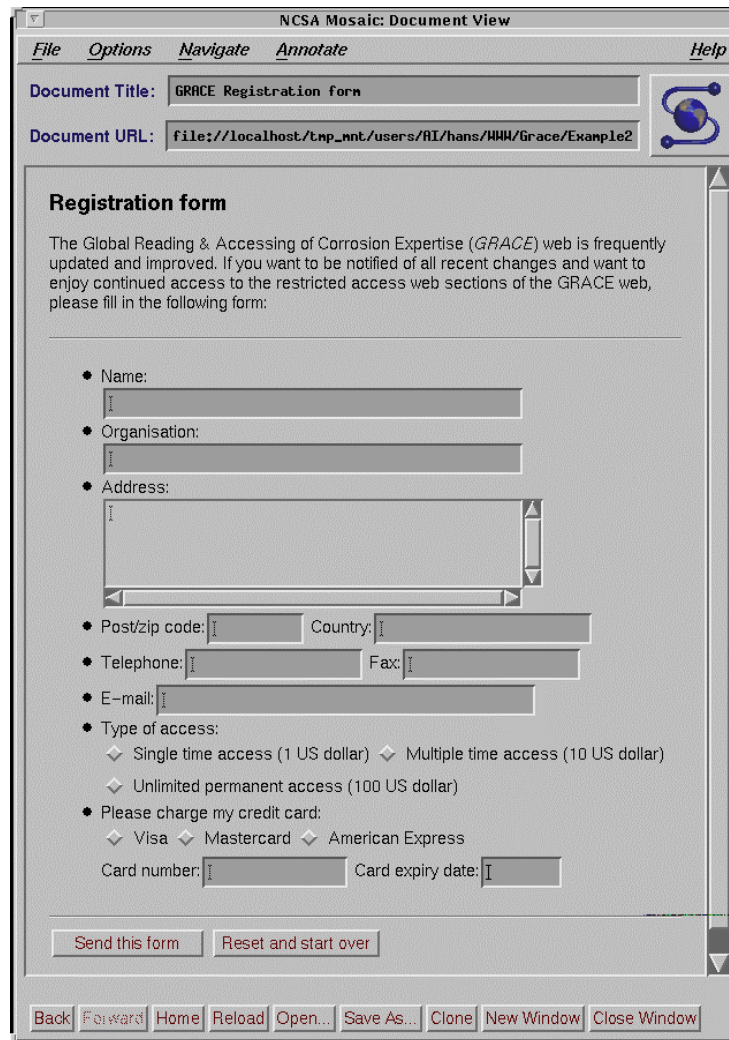
HTML - Example: a form

```
<html>
<head>
<title>GRACE Registration form</title>
<link rev="made" href="mailto:hca@mtm.kuleuven.ac.be">
</head>
<body>
<h1>Registration form</h1>
<p>The Global Reading &amp; Accessing of Corrosion Expertise (<em>GRACE</em>) web is frequently updated and improved. If you
  want to enjoy continued access to the restricted access web sections of the GRACE web, please fill in the following form:</p>
<form method="POST" action="file://localhost/WWW/GRACE/cgi-bin/register.pl"><hr>
<ul>
<li>Name:<br><input type="text" size=50 name="Name"></li>
<li>Organisation:<br><input type="text" size=50 name="Organisation"></li>
<li>Address:<br><input type="text" size=50,5 name="Address"></li>
<li>Post/zip code:<input type="text" size=10 name="Code">Country:<input type="text" size=25 name="Country"></li>
<li>Telephone:<input type="text" size=20 name="Phone">Fax:<input type="text" size=20 name="Fax"></li>
<li>E-mail:<input type="text" size=45 name="Email"></li>
<li>Type of access:<br>
  <input type="radio" name="access" value="Single">
  Single time access (1 US dollar)
  <input type="radio" name="access" value="Multiple">
  Multiple time access (10 US dollar)
  <input type="radio" name="access" value="Unlimited">
  Unlimited permanent access (100 US dollar)</li>
<li>Please charge my credit card:<br>
  <input type="radio" name="cardtype" value="Visa">Visa
  <input type="radio" name="cardtype" value="Mastercard">Mastercard
  <input type="radio" name="cardtype" value="American Express">American Express<br>
  Card number:<input type="number" name="cardnumber" size=16>
  Card expiry date:<input type="number" name="carddate" size=8></li>
</ul><hr>
<input type=submit value="  Send this form  ">
<input type=reset value=" Reset and start over ">
</form>
</body>
</html>
```



How the WWW works

HTML - Example: a form



The screenshot shows a window titled "NCSA Mosaic: Document View". The menu bar includes "File", "Options", "Navigate", "Annotate", and "Help". The "Document Title" is "GRACE Registration form" and the "Document URL" is "file://localhost/tmp_nnt/users/RL/hans/WWW/Grace/Example2".

Registration form

The Global Reading & Accessing of Corrosion Expertise (*GRACE*) web is frequently updated and improved. If you want to be notified of all recent changes and want to enjoy continued access to the restricted access web sections of the GRACE web, please fill in the following form:

- Name:
- Organisation:
- Address:
- Post/zip code: Country:
- Telephone: Fax:
- E-mail:
- Type of access:
 - ◇ Single time access (1 US dollar) ◇ Multiple time access (10 US dollar)
 - ◇ Unlimited permanent access (100 US dollar)
- Please charge my credit card:
 - ◇ Visa ◇ Mastercard ◇ American Express
- Card number: Card expiry date:

Buttons: "Send this form" and "Reset and start over".

Navigation buttons: "Back", "Forward", "Home", "Reload", "Open...", "Save As...", "Clone", "New Window", "Close Window".

How the WWW works

HTML - Consecutive generations

HTML 1.0

used in 1994

- “what is parsed by Mosaic” (looks like SGML)

HTML 2.0

widely used now

- standardized DTD (fully conformant SGML)
- fill-in forms

HTML 3.0

under development

- tables and better formatting

HTML 4.0

ready for use in 1996?

- maths and figures



How the WWW works

HTTP: HyperText Transfer Protocol

Protocol for the transmission of HTML documents

- stateless
- object-oriented
- fast and lightweight

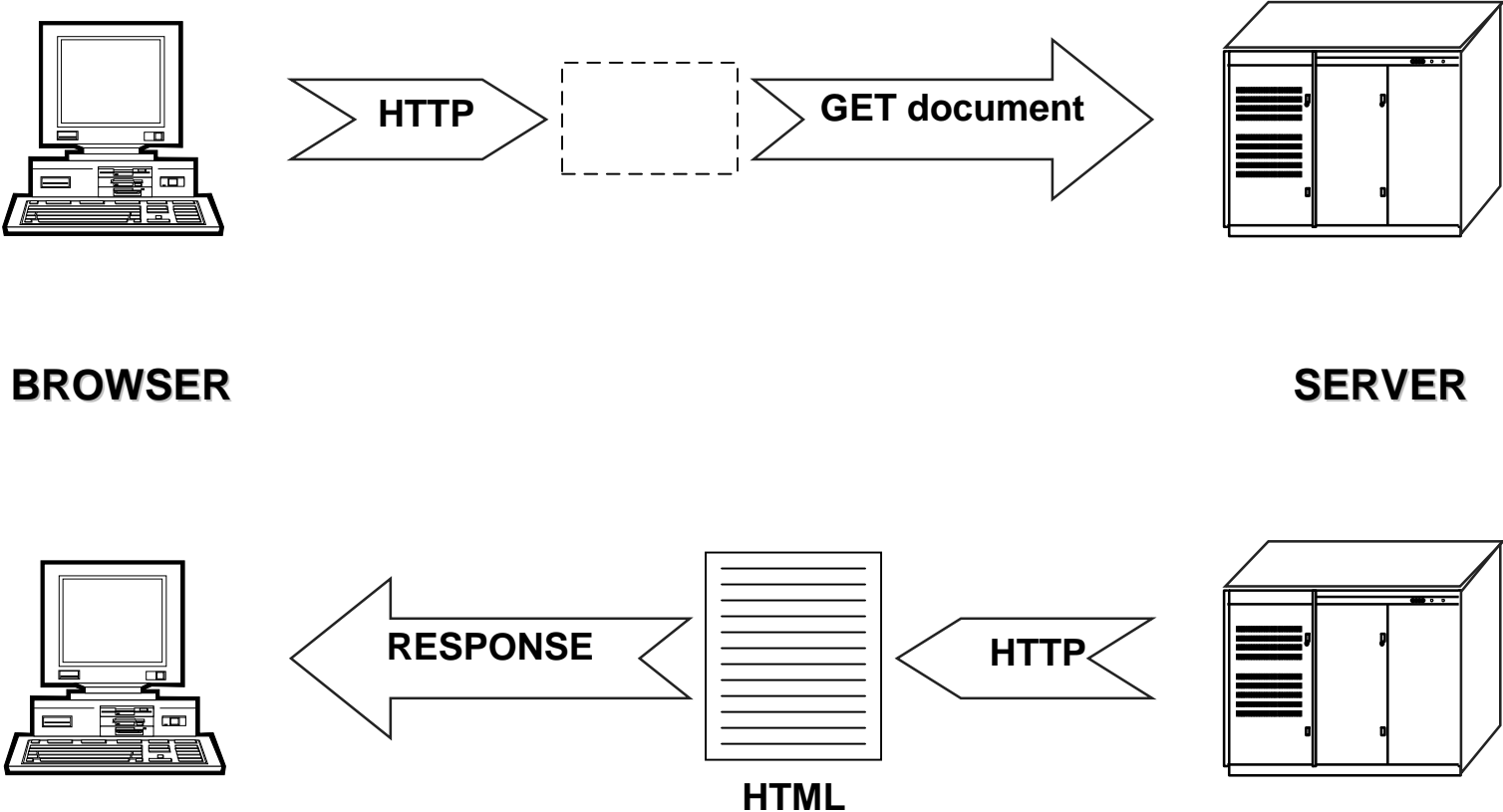
Transaction established over a TCP/IP connection:

- *connect* establishment of a connection between browser and server
- *request* sending by the browser of a request message to the server
- *response* sending by the server of a response message to the browser
- *close* closing down of the connection by the browser or the server



How the WWW works

HTTP - Transaction



How the WWW works

HTTP - Transaction steps

REQUEST message: Method URI Header Data

- Method what to do with the object (e.g. GET, PUT, HEAD, ...)
- URI where to find the object (URL)
- Header what is the object (HTTP header, MIME header)
- Data the object itself (ASCII text, raw binary data, ...)

RESPONSE message: StatusCode ReasonLine

- StatusCode a status code number for the browser
- ReasonLine an explanation for the human reader

e.g. 404 Not found, 502 Service temporarily overloaded



How the WWW works

MIME: Multimedia Internet Mail Extension

Standard for encoding multimedia data (RFC 1341)

- ASCII, HTML, PostScript, ...
- GIF, JPEG, MPEG, QuickTime ...

Format negotiation

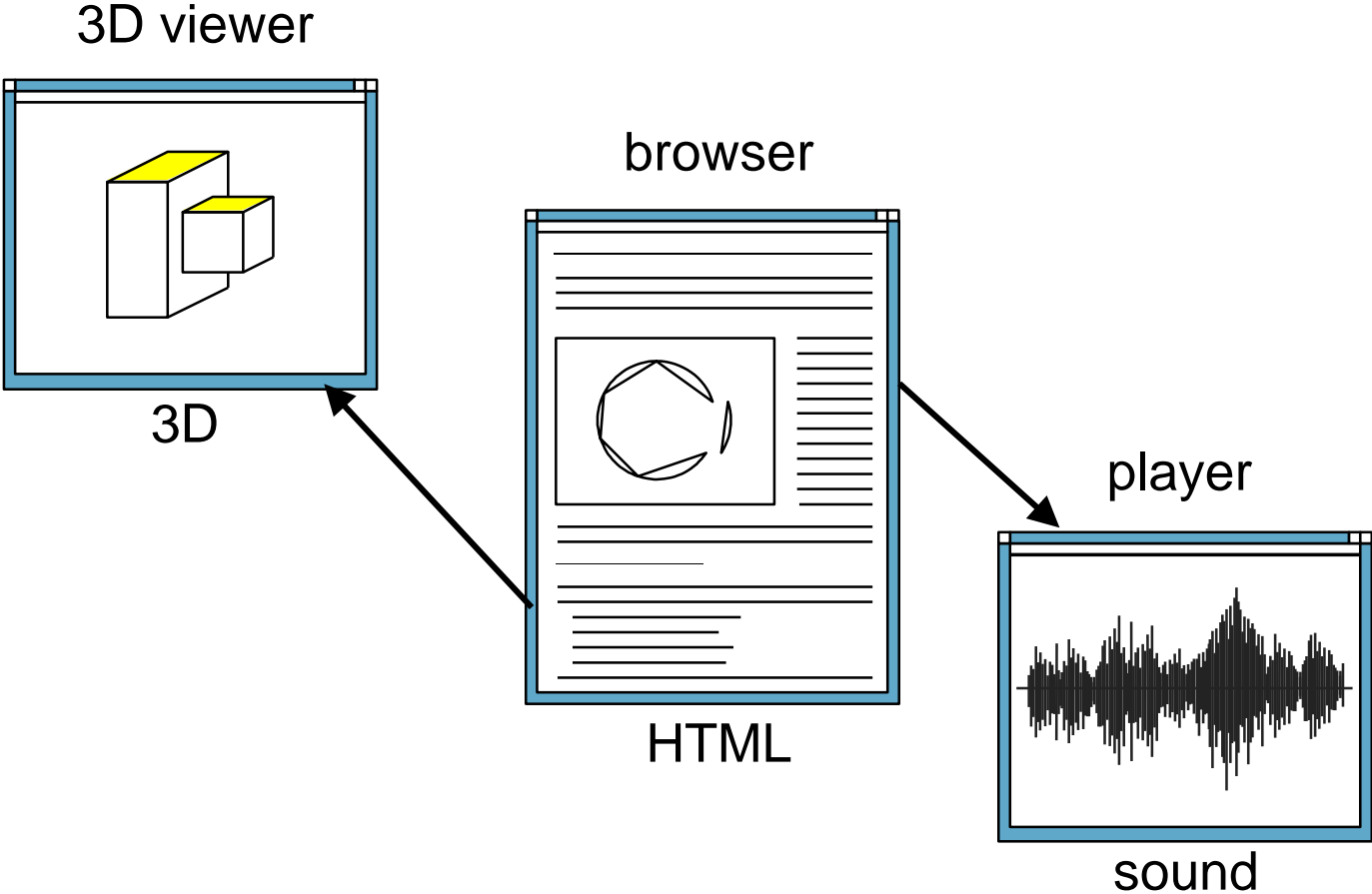
- server notifies browser of available data formats
- browser notifies server of acceptable data formats
- if necessary, server converts to the appropriate data format
- server sends browser the appropriate data format



How the WWW works

MIME - Browser extensibility

Browser + helper applications



How the WWW works

CGI: Common Gateway Interface

An interface for *gateways* (external programs)

- that handle information requests or queries coming from a browser
- that return the appropriate data or generate a HTML document on the fly

Possible uses:

- accessing legacy databases and returning HTML results
- allowing user feedback to the server through a fill-in HTML form

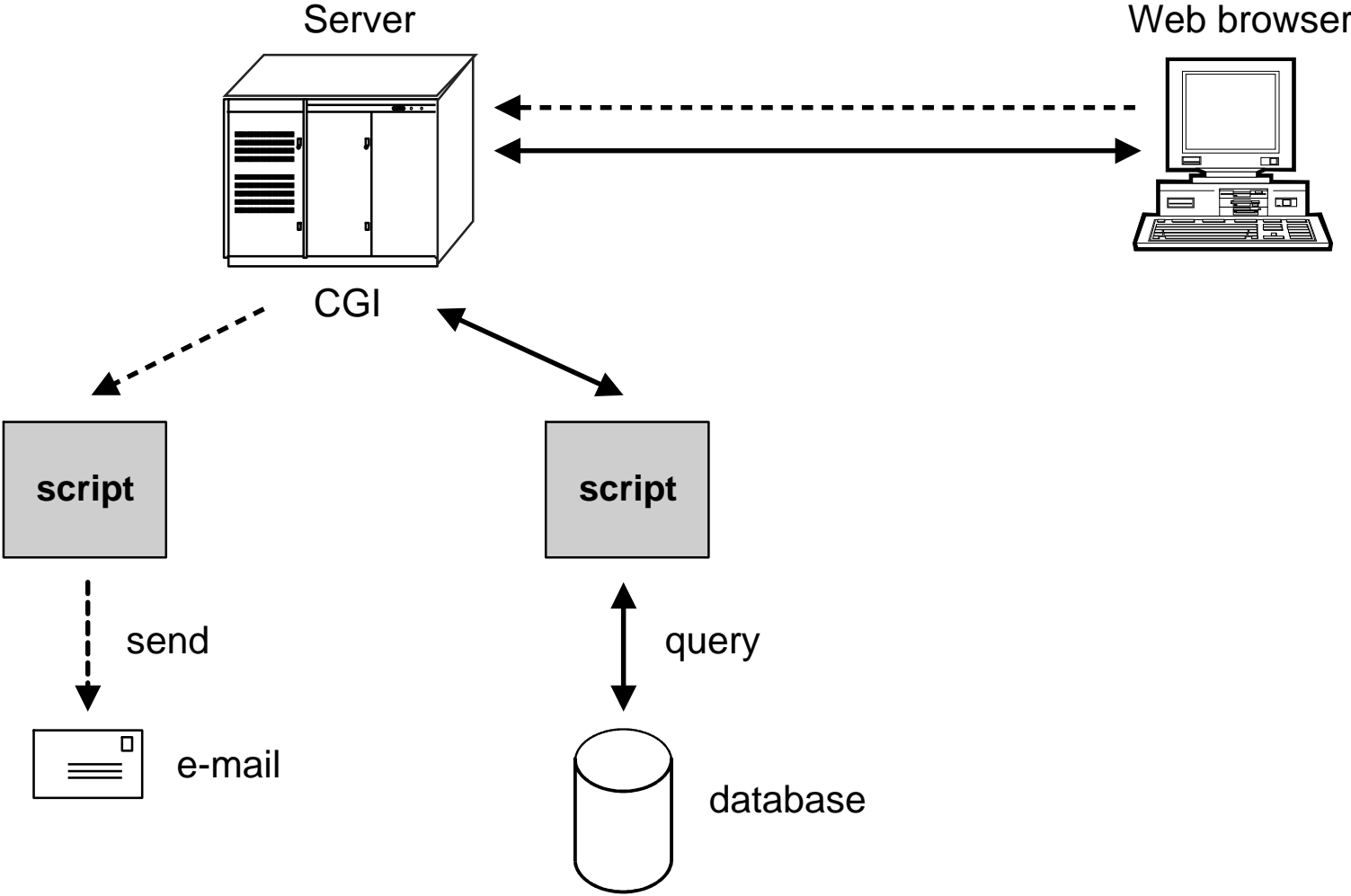
Languages:

- C/C++, Perl, TCL, Bourne Shell, C Shell, Lisp, Prolog, ...



How the WWW works

CGI - Server extensibility



How the WWW works

Web robots and web extensions

Webcrawlers and webspiders:

- collect data on web size and growth
- maintain distributed hypertext infostructures
- collect or extract indexing terms for use in web indices

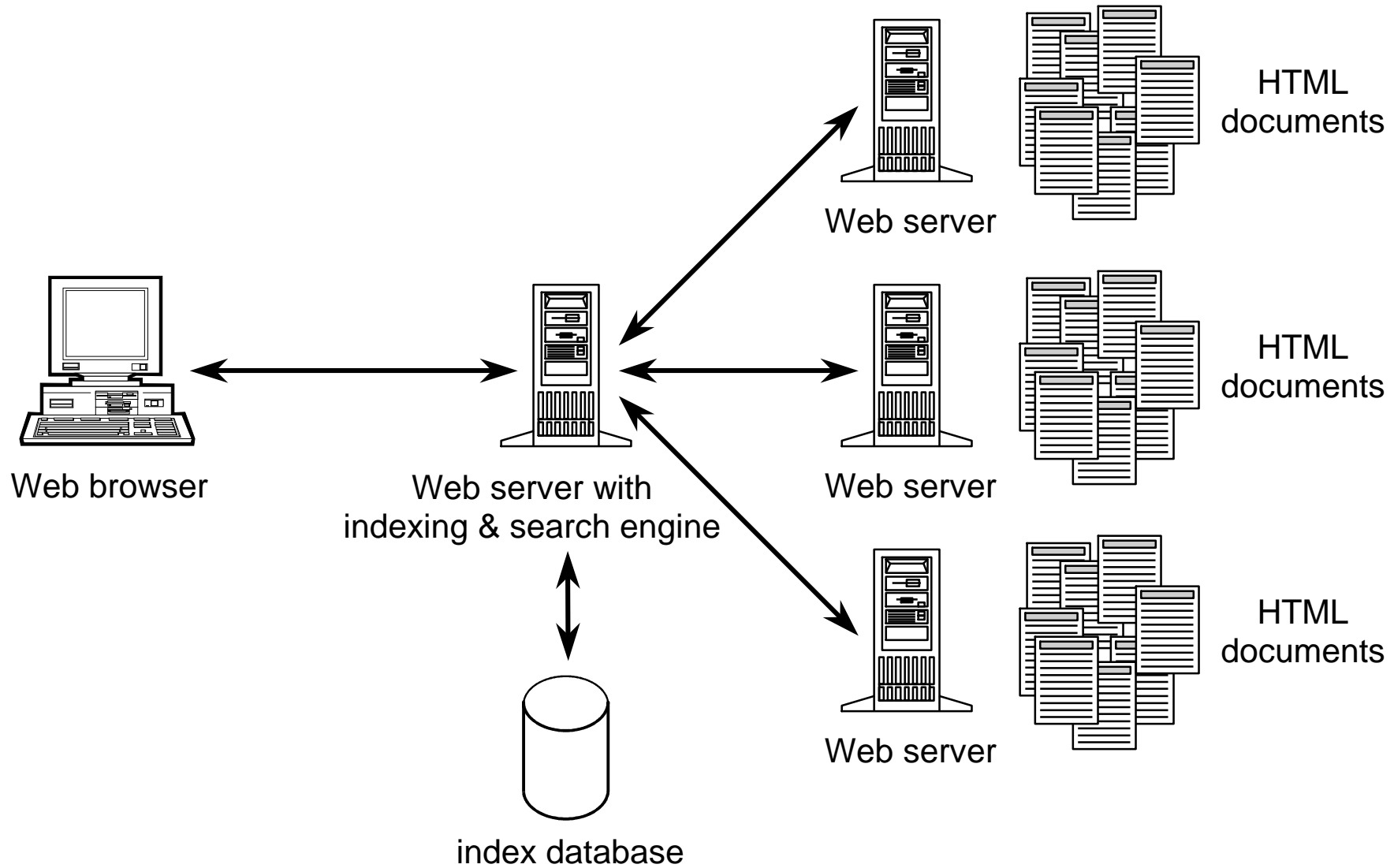
Web caches, web indices and web brokers:

- keep local copies, refreshed on a regular basis
- serve as overview maps or starting points for navigation
- act as agents searching for documents of particular interest



How the WWW works

Web robots and web extensions



How the WWW works

Web robots and web extensions

Webcrawlers and webspiders:

- WebCrawler <http://webcrawler.cs.washington.edu/WebCrawler/>
- WWWworm <http://www.cs.colorado.edu/home/mcbryan/WWW.html>
- MOMspider <http://www.ics.uci.edu/WebSoft/MOMspider/>

Web caches and web indices:

- Lagoon <http://www.win.tue.nl/lagoon/>
- Harvest <http://rd.cs.colorado.edu/harvest/>
- JumpStation <http://www.stir.ac.uk/jsbin/js/>



How the WWW works

Future web developments

***Secure HTTP* : HTTP for a Secure Mosaic**

- extension of HTTP: standard message format and protocol
- support for different encryption algorithms and operation modes
- client authentication, transaction data encryption, billing data encryption

***HSSSL* : Hypertext Style and Semantics Specification Language**

- separate presentation from content
- stylesheets with styles for HTML elements

***VRML* : Virtual Reality Markup Language**

- markup language for 3D interfaces



How to use the WWW

Living in the global infospace: why use it?

What to look for:

- software products: public domain, shareware, bug fixes, ...
- research results: experimental data, technical reports, papers, ...
- marketing, product, financial information to raise competitive awareness

How to find it:

- consult a commercial infoservice which targets your needs
- query or browse through the entries of a web index
- find a good homepage and start from there




How to use the WWW

Living in the global infospace: why be on it?

POP : Point of Presence

- establish an on-line presence as part of hi-tech corporate image
- provide access to company-specific information (products, jobs, ...)

POA : Point of Access

- market real-world products or services (e.g. catalogue, museum, ...)
- deliver web support services: editing/indexing/filtering  *infoware*

POS : Point of Sales

- sell products: consumer products, industrial equipment, ...
- sell information: journals, technical reports, financial data, entertainment, ...



How to use the WWW

Living in the global infospace: why be on it?

POP : Point of Presence

• government <http://www.whitehouse.gov/>, <http://www.who.ch/>

• industry <http://www.microsoft.com/>, <http://www.ibm.com/>

POA : Point of Access

• online exhibits <http://mistral.enst.fr/louvre/>, <http://kspace.com/>

• resource guides <http://www.ora.com/gnn/>, <http://www.digimark.net/wow/>

POS : Point of Sales

• products <http://www.pizzahut.com/>, <http://www.internet.net/>

• services <http://www.quote.com/>, <http://www.espan.com/>

• magazines <http://www.timeinc.com/>, <http://www.ziff.com/>



How to use the WWW

Living in the global infospace: why not be on it?

No mature infomarket, but ...

- size and diversity of users (information superhighway)
- different selling strategies (subscription, on-demand)

No reliable tools and technology, but ...

- shrink-wrapped software (commercial browsers)
- infinite bandwidth to the home (cable, ISDN, ATM)

No appropriate commercial infrastructure, but ...

- transparent encryption (public keys, SOAPs)
- reliable billing mechanism (eDollar, digital cash)



How to use the WWW

Setting up a local infospace: infrastructure

Network demands

- TCP/IP on top of existing network
- ISDN dial-up or permanent connection

Server administration

- hire a full-time webmaster
- don't underestimate server load

Browser management

- maintain access control
- "latest version" is moving target



How to use the WWW

Setting up a local infospace: infostructure

Document creation

- conversion RTF, FrameMaker, LATEX, ... à HTML
- downtranslation SGML à HTML (e.g. EBT's DynaWeb)

Document design

- get the reader's attention first!
- visual consistency and link reliability

Navigation support

- try to avoid "getting lost in hyperspace"
- provide orientation cues, landmarks, search tools, ...



Summary and conclusions

Where we will jump to next ...

World-Wide Web

- the **Wild West** is **Waiting** ...
- the new frontier for software and services
- coming soon to a *Flemish TeleNet* socket near you?

Jump in and see for yourself ...

- industry <http://www.bekaert.com/>
- banking <http://rabobank.info.nl/>
- publications <http://www.riv.nl/>
- entertainment <http://www.vpro.nl/>

