

# WWW - Invention (hypertext)

- **Invented by Tim Berners-Lee at CERN in 89**
  - "HyperText is a way to link and access information of various kinds as a web of notes in which the user can browse at will. It provides a single user-interface to large classes of information."
  - "A hypertext page has pieces of text which refer to other texts. Such references are highlighted and can be selected... When you select a reference, the browser presents you with the text which is referenced: you have made the browser follow a hypertext link"

2/24/04

2

## WWW - Invention

- **Tim Berners-Lee (CERN) Proposal 11/90**
  - "The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, ..."
  - "A link is specified as an ASCII string from which the browser can deduce a suitable method of contacting an appropriate server. When a link is followed, the browser addresses the request for the node [document] to the server."
  - WorldWideWeb: Proposal for a HyperText Project, 11/12/90, <http://www.w3.org/Proposal.html>

# Links may become invalid

- Link is simply a text name for a remote document
- Remote document may be removed while name in link remains in place
- Guys who invented hypertext worried about this (Douglas Engelbart invented and implemented hypertext in early 60s, Ted Nelson named it in '65, continues pursuing 'Xanadu')
- Berners-Lee had brilliance to ignore it

## Uniform Resource Identifiers (URI)

- An extensible scheme for identifying resources
  - Uniform - common method for naming, locating things
  - Resource - any entity (page, server, human)
  - Identifier - character string that identifies the entity
- *URL* are a subset of URI that identify resources by their primary access mechanism (HTTP, ftp, etc).
- *URN* are a subset of URI that identify resources that are globally unique and persist even when resource disappears.
- **RFC 2396**

# Identifying a Page (URL) RFC 2396

- Page identified by:
  - Protocol used to access page
  - Computer on which page is stored
  - TCP port to access page
  - Pathname of file on server
- URL Syntax

*protocol://computer\_name[:port][/document\_name]*

- *protocol* (scheme) = http, ftp, etc
- *port* is optional
- *document\_name* is path to document
- Which parts are case sensitive?

## URL Schemes

```
ftp : // [ user [:password] @ ] host / path
news : newsgroup
telnet : ipaddress
gopher : // host [ gtype ]
mailto : userid @ hostname
wais : // hostport / database [ ? search ]
wais : // hostport / database / wtype / wpath
file: // pathname
http: // host [: port ][ / path ]
```

- <http://www.w3.org/Addressing/rfc1738.txt>

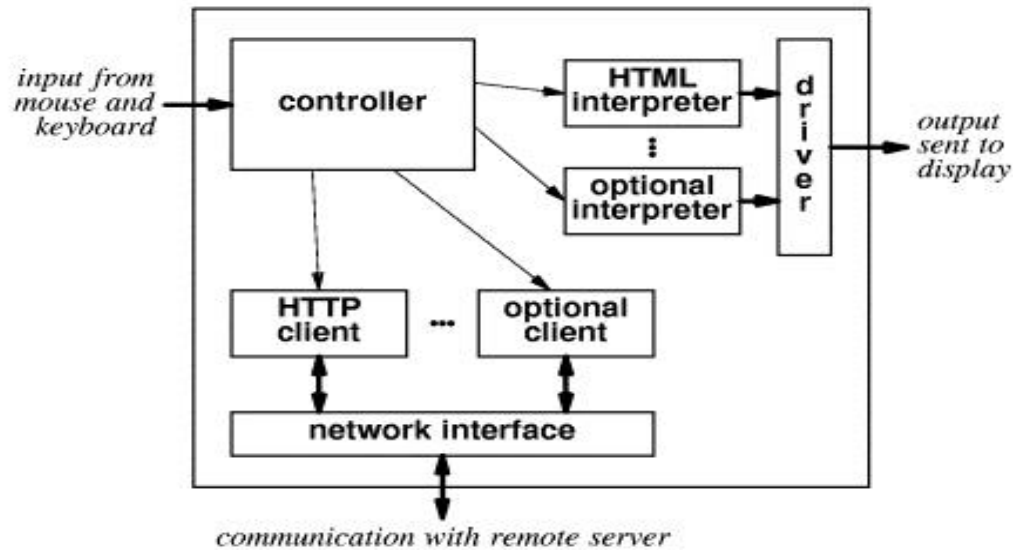
# How Did Berners-Lee Start the Web Growing?

- What good were Web servers if there were no browsers?
- What good were browsers if there were no Web servers?
- Why were any of them deployed?

## Partial (Relative) URI's

- Purpose
  - Allow relative reference among objects in the same hierarchy
  - Enable relocation of a set of objects in a hierarchy
- Enabled by hierarchical delimiters
  - / . . (slash, dot dot, dot)

# Browser Architecture



## Browser Functionality

- Network access
  - Protocol implementations
    - HTTP client to fetch documents from WWW servers
    - Clients for other protocols (e.g., ftp, news, etc.)
  - Connection management
  - Caching
- GUI Display
  - Multiple document representation interpreters
    - HTML interpreter for HTML-formatted documents
    - Text
    - Plug-ins for other interpreters (e.g., Shockwave, Adobe) for other media types
  - Complex layout, such as frames