

## Il Servizio World Wide Web (WWW)

Corso di *Informatica Generale*  
(Roberto BASILI)

Teramo, 20 Gennaio, 2000

## Il World Wide Web

- Storia
- WWW: obiettivi
- WWW: tecnologie
- Le Applicazioni
- Scenari Futuri

### WWW

- La Storia
  - HTTP (1990)
  - Gopher (1991)
  - Mosaic (CERN, 1993)
  - Netscape (inizio 1994)

### WWW

- Il World Wide Web e' un *sistema distribuito* per lo scambio di **informazione multimediale**.
- Il sistema distribuito si basa su un paradigma non lineare di documentazione: **l'ipertesto**

### WWW

- Per informazione multimediale si intende un insieme di dati riferibili a differenti tipologie di comunicazione:
  - testuale e numerica (alfanumerica)
  - vocale
  - sonora
  - visuale
- L'aggregazione significativa di tali livelli di rappresentazione forma il documento multimediale

### WWW

- Nel parlare di multimedialita' si deve distinguere
  - eterogeneita' dei supporti di comunicazione
  - eterogeneita' delle forme di comunicazione, e quindi
  - eterogeneita' dei linguaggi di rappresentazione

## WWW

- Un ipertesto e' un documento che, abolendo la struttura lineare propria delle forme tradizionali del linguaggio, supporta forme libere di accesso (non lineari)
- Tale *non linearita'* genera una organizzazione *reticolare* dell'informazione
  - le *unita' informative* sono i **nodi** della rete (*ancore*)
  - i **collegamenti tra i nodi** (*link*) possono essere attraversati liberamente durante la "lettura"

## WWW

- Quando le unita' informative fanno riferimento ad informazione veicolata da linguaggi diversi (per es. suono e testo, testo e immagini) l'ipertesto si generalizza alla nozione di *ipermedia*

## WWW

La non linearita' influenza profondamente

- il *processo di scrittura* detto usualmente "**authoring**" dell'ipertesto
- il *processo di fruizione* che avviene mediante una proattivit  dell'utente) (cioe' **interattivamente**)

La "lettura" e' usualmente detta "**navigazione**" (o "*browsing*")

## WWW

- Tecnicamente:
  - i nodi informativi sono nuclei di informazioni automatizzate (per es. files)
  - i collegamenti sono legati ad azioni (per es. pulsanti) agenti su tali nuclei che attivano processo di comunicazioni tra elaboratori in rete
- L'automazione di tali processi e' basata su architetture di rete client-server

## WWW

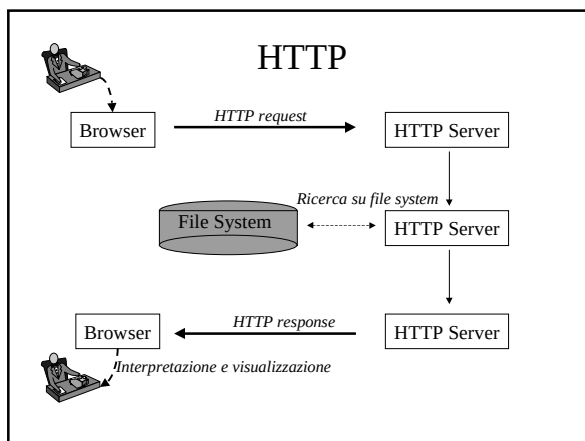
- Il WWW (un sistema distribuito per la comunicazione ipertestuale di informazione multimediale organizzata in forme ipertestuali) e' quindi un servizio di Internet
- Usa:
  - HTTP come protocollo di trasmissione
  - HTML come protocollo di rappresentazione ipertestuale

## WWW

- I Protocolli di WWW
  - HTTP
    - The Hypertext Transfer Protocol (HTTP) is an **application-level protocol** for distributed, collaborative, hypermedia information systems
  - HTML (Hypertext Markup Language)
    - The HyperText Markup Language (HTML) is a simple **data format** used to **create** hypertext documents that are portable from one platform to another.

## HTTP

- HTTP definisce il tipo e la negoziazione di un dato da comunicare in modo da consentire alle applicazioni che ne fanno uso di essere indipendenti dai dati da trasferire
- Definisce flessibilmente
  - Lo scopo di una richiesta
  - Informazione di localizzazione
  - Il formato dei dati di una richiesta
  - ... E della risposta
  - Modifiche da apportare



## HTTP

- Terminologia
  - HTTP client
  - HTTP server
  - messaggio
  - richiesta/risposta
  - negoziazione dei contenuti (MIME)
  - user agent (e.g. browsers, spiders, ...)
  - tempo di terminazione (expiration time)

## HTTP

- Localizzazione della informazione
- I nomi logici delle entità informative nella rete (files su host) sono specificate da una convenzione che dipende
  - dal protocollo usato per accedervi correttamente (per es. FTP, GOPHER, HTTP)
  - dall'host che li ospita
  - dalla loro localizzazione nel file system

## HTTP

- Universal Resource Identifier (URI) e' il nome univoco che una entità (membro dello spazio dei nomi) possiede
  - SCHEMA:PATH[#Etichetta]
- SCHEMA, individua il protocollo
- PATH specifica dominio, host e file
- Etichetta specifica il frammento del documento

## HTTP

- Attualmente e' in uso un sottoinsieme di URI denominato URL (Uniform Resource Locator), che si riduce a  
**SCHM://[user:passwd@]HOSTID[:PORTA]/[PATH][#Et]**
- es.  
**http://www.uniroma2.it:80/home/index.html**  
**file:///C:/Basil/Pres/WWW/Terms.html#node**

## Documenti Web

- Documenti Web statici
  - sono costituiti di insiemi di informazioni multimediali
  - ... generalmente memorizzate in files
  - Il server HTTP riceve la loro richiesta
  - ... li localizza sul proprio file system
  - ... ne invia una copia al client

## Documenti Web

- Documenti Dinamici
  - esistono solo “durante” una interazione Server-client HTTP
  - vengono generati quando e’ necessario costruire i contenuti a fronte di una interazione utente (es. query ad Altavista) o di una richiesta (dinamicamente)

## Documenti Web

- Documenti Dinamici
  - la costruzione non e’ ad opera del client (user agent) o del server HTTP ma da programmi (o librerie) esterne
    - Common Gateway Interface (CGI)
    - ASP (Active Server Page) (Explorer)
    - ISAPI/Java (Netscape)
    - Perl (Apache)

## Documenti Web

- CGI
  - stabilisce regole per lo scambio di informazioni tra il browser B, il server HTTP H, ed il programma chiamante P
  - B produce le informazioni della richiesta secondo il formato CGI
  - H decodifica ed invia le informazioni a P
  - P restituisce le informazioni (secondo lo stile HTML) ad H
  - H invia a B la risposta

## HTML

- E’ un linguaggio per la descrizione dei documenti ipertestuali che sono disponibili nel WWW
- E’ orientato alla descrizione dei formati ed e’ poco sensibile ai contenuti
- Possiede un insieme di estensioni che consentono
  - il trattamento di dati multimediali
  - la gestione dei collegamenti ipertestuali

## HTML

- E’ un linguaggio di *mark-up* (marcatura) perche’ rappresenta l’informazione (formato ed estensioni) mediante etichette (*markers* o *tags*) che vengono aggiunte al tradizionale contenuto testuale
- Es.
  - `<H1>Hypertext Terms</H1>`
  - `<B>Hypertext Terms</B>`
  - `<A NAME=Web>Web Section</A> >`

## HTML

- E' giunto alla versione 4.0 (Dicembre 1997)
  - uso dei frames
  - uso dei fogli di stile
- un documento HTML e' un documento di testo (ascii) che grazie al mark-up
  - viene interpretato dal browser
  - e visualizzato in accordo con i significati che ai TAG vengono assegnati dal browser

## HTML

- La gestione di informazione multimediale (inclusione di files non testuale) e' gestita da programmi esterni al browser (estensioni) che secondo gli standard MIME ne garantiscono la corretta visualizzazione
- il browser consente alla estensione (in genere una libreria dinamica) di accedere al file e di costruire la corretta visualizzazione