



Internet ed il World Wide Web

UNIVERSITA' DI MACERATA
Corso di laurea triennale in Formazione e
Management dei sistemi turistici
3 CFU

II periodo

Barbara Re
barbara.re@unicam.it

Internet ed il World Wide Web



- Il servizio WEB fornisce la possibilità di realizzare un **documento ipertestuale** costituito da **pagine** memorizzate su differenti nodi della rete internet
- **WEB** = ragnatela
- **Documento ipertestuale** = documento con link (collegamenti) ad altre parti di documenti

Web come Mezzo di Comunicazione



- La pagine del servizio WEB permettono spesso l'accesso ad altri servizi
 - ✘ il Web può essere visto quindi come un modo per accedere a servizi Internet in generale (videoconferenze, trasmissioni audio, chat, ...)
- Per questo motivo il WEB
 - ✘ è divenuto il servizio INTERNET più diffuso, tanto che spesso i due termini vengono confusi
 - ✘ è divenuto un mezzo di comunicazione di importanza paragonabile a quella di Radio, TV e telefono

3

Caratteristiche del Web



- Possiede sia caratteristiche
 - ✘ di *broadcast* (diffusione), come TV e radio (esistono già radio presenti accessibili via Web)
 - ✘ che *interattive*, come il telefono (è già possibile fare telefonate tramite il Web)
- Vantaggi: *maggior democrazia*
 - ✘ chiunque può diffondere (pubblicare) le proprie pagine
- Svantaggi: *sviluppo incontrollato*
 - pagine spazzatura e difficoltà di reperimento informazioni

4

Un po' di storia



- Il progenitore del WEB moderno è stato un text-based Web browser scritto nel 1990 da **Berners-Lee** (un ricercatore del European Particle Physics Laboratory - CERN)
- **Mosaic** è il primo graphical Web browser e venne scritto nel 1993 dal NCSA (National Center for Supercomputing Applications)

5

Il successo del Web



- Calcolatori potenti e poco costosi
- Diffuso accesso a Internet
- Un linguaggio semplice da usare (HTML)
- Combinazione con grafica, suoni, filmati (multimedialità)
- Combinazione con altri servizi Internet
- Browser semplici da ottenere ed utilizzare
- Grande attenzione dei mass-media

6



- La tecnologia legata al Web viene discussa e gestita all'interno di una organizzazione internazionale chiamata **W3C** (World Wide Web Consortium) che comprende:
 - Organismi governativi
 - Centri di ricerca (pubblici, privati, universitari)
 - Realtà produttive (aziende, industrie, ...)



- Uno dei problemi principali legati al Web riguarda l'identificazione delle risorse disponibili
- Il W3C ha introdotto dei particolari identificatori detti: **Uniform Resource Identifier (URI)**
- Esistono due tipi di URI
 - **Uniform Resource Locator (URL)** identifica una risorsa indicando l'indirizzo presso cui reperirla
 - **Uniform Resource Name (URN)** identifica una risorsa persistente (che rimarrà sempre disponibile allo stesso indirizzo) o mobile (il cui indirizzo può cambiare)

Come interpretare un URL



- Gli URL consistono di 3 parti:
 - il nome del **protocollo** da utilizzare per l'accesso alla risorsa (es. http, ftp, mailto, ...)
 - il nome della macchina **host** della risorsa (si può usare sia l'IP address che il nome logico)
 - il **percorso** ed il **nome** della risorsa (detto **pathname**)

Esempi:

<http://webs.students.dsc.unibo.it/calendario.html>

ftp://cs.unibo.it/pub/techreports/2000_01.ps.gz

<mailto:zavattar@cs.unibo.it>

9

Browser



- I browser sono programmi applicativi che
 - dato un URL
 - utilizzano il protocollo specificato dall'URL per caricare la risorsa indicata
 - la decodificano e la **rendono** (visualizzano, traduzione di **rendering**)
- Esempi:
 - Internet Explorer, Netscape Navigator

10

Reperire informazioni sul Web



- Il modo più semplice di navigare il web è:
 - ✘ indicare un URL di partenza
 - ✘ seguire i link disponibili
- Per cercare informazioni senza conoscere URL in cui reperirle si possono usare due tipi di strumenti:
 - **WEB SEARCH ENGINE** (motori di ricerca)
 - **SUBJECT TREES** (indice gerarchico di argomenti)

11

Web Search Engine



- I motori di ricerca si basano su database di pagine indicizzate
 - Ogni URL conosciuta viene archiviata con alcune parole chiavi associate
 - **Full text indexing**: tutte le parole presenti nella pagina vengono considerate parole chiave
 - **Selective text indexing**: solo le parole ritenute più importanti vengono considerate parole chiave

12

Spider



- Le pagine da indicizzare vengono visitate dai motori di ricerca tramite strumenti detti **spider**
 - ✘ Visitano il web raccogliendo URL e relative parole chiavi
- Problemi:
 - Non è possibile **visitare l'intero Web** a causa della sua estensione e del suo continuo mutare
 - Non è possibile **mantenere aggiornata** la visione del Web da parte dei motori di ricerca a causa del fatto che queste possono essere modificate o rimosse

13

Search Query



- Per attivare la ricerca si deve formulare una **search query** (interrogazione di ricerca)
- Il risultato è un elenco di **hit** (URL di pagine che soddisfano la richiesta)
- Le hit vengono elencate in ordine di rilevanza rispetto alla search query. Vengono usate diversi **metriche**:
 - ✘ Pagine che contengono più volte le parole chiave
 - ✘ Pagine su cui esistono più link in generale
 - ✘ Pagine su cui esistono più link all'interno delle hit considerate
- **Noise** (Rumore): fenomeno a causa del quale buona parte di hit non sono interessanti
- Esempi di motori di ricerca: **altavista, google, lycos**

14

Subject Tree



- Le pagine indicizzate sono strutturate per argomenti
- Gli argomenti sono organizzati secondo un albero gerarchico
- L'albero può essere visitato partendo dagli argomenti più generali che vengono via via specializzati in sottoargomenti scendendo la gerarchia
- In questi casi, oltre agli spider i progettisti di pagine web possono sottomettere le proprie pagine indicando gli argomenti relativi
- Esempio: **yahoo**, **arianna**

InfGen(3) - 15⁵

Portali



- I search engine e subject tree sono oramai stati incorporati nel più generale concetto di **PORTALE**
- Un portale è un punto di accesso ad internet che fornisce vari tipi di servizi:
 - ✗ Accesso gratuito ad Internet
 - ✗ Caselle di posta e Spazio per pagine web gratuiti
 - ✗ Search Engine e Subject Tree
 - ✗ Informazioni aggiornate di carattere generale
- Un portale specializzato per uno specifico argomento viene detto **VORTAL** (portale di carattere verticale)

16



- Le singole pagine WEB sono rappresentate attraverso un linguaggio chiamato HTML

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN">
<HTML>
  <HEAD>
    <TITLE>Prova</TITLE>
  </HEAD>
  <BODY>
    La mia prima pagina web
  </BODY>
</HTML>
```



- Il **Markup** consiste di informazioni aggiunte inserite in un testo per applicarvi effetti
- Permette una comunicazione più chiara, più leggibile, più invitante dell'informazione contenuta nel testo
- Esempi:
 - formattazione (grassetto, sottolineato,...)
 - punteggiatura
 - spazi

Tipi di Markup



- **Puntuazionale:** segni che apportano informazioni sintattiche (es. ; , : .)
(N.B. esistono ambiguità: fine frase, abbreviazioni, sospensione)
- **Presentazionale:** effetti grafici per migliorare la presentazione (es. paragrafi, punti elenco)
- **Procedurale:** indicazione di meccanismi di presentazione da passare a strumenti di visualizzazione (esempio: effetti powerPoint)
- **Descrittivo:** definizione di ruoli all'interno del testo (titolo, paragrafo,...)
- **Referenziale:** fare riferimento ad entità esterne al documento (es. collegamenti ad allegati visivi o sonori)

19

HyperText Markup Language



- Le pagine Web vengono codificate in formato testo arricchito con informazioni di markup (dette **direttive di markup**)
- Le direttive di markup vengono inserite nel testo tramite **TAG**:

HTML = testo + TAG

es: <P>Questo è un paragrafo.</P>
<H1>Titolo di primo livello</H1>

20

La sintassi delle direttive



- Ogni direttiva di markup ha:
 - un **nome**, es. P H1 H2 TITLE
 - un tag di **apertura** ed uno di **chiusura**
apertura: <P> chiusura: </P>
 - i tag di apertura possono contenere eventuali attributi opzionali
<H1 align="center">Titolo centrato</H1>
 - align nome attributo
 - center valore

21

Struttura di un documento HTML



- Il documento inizia con la direttiva DOCTYPE:
* <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN">
- Il resto del documento è racchiuso fra:
<HTML> </HTML>
- Questa parte si divide in 2 sezioni:
 - Le specifiche di testa:
<HEAD> </HEAD>
 - Il corpo:
<BODY> </BODY>

22

Le specifiche di testa



- L'elemento **TITLE**: serve per specificare il titolo del proprio documento
`<TITLE>Visualizzato nella barra del titolo</TITLE>`
- L'elemento **META**: serve per inserire informazioni aggiuntive
`<META http-equiv="refresh" content="180,http:...">`
- L'elemento **LINK**:
`<LINK href="b.css" rel="stylesheet" type="text/css">`

23

Il corpo del documento



- Il corpo del documento contiene ciò che verrà visualizzato
`<BODY bgcolor="#008800"> </BODY>`
- Il testo verrà **reso** solo in funzione delle direttive di markup
- Per inserire testo da non visualizzare viene inserito fra:
`<!-- Questo è un commento -->`

24

Caratteri Speciali



- Alcuni caratteri non possono essere inseriti direttamente da tastiera, ma bisogna usare dei codici associati:

<code>&lt;</code>	<	<code>&egrave;</code>	è
<code>&gt;</code>	>	<code>&eacute;</code>	é
<code>&amp;</code>	&	<code>&igrave;</code>	ì
<code>&quot;</code>	"	<code>&ograve;</code>	ò

- Altri tipi di accenti:
circ, ring, tilde, dieresi, cedilla

25

Strutturazione di blocchi



- Esistono diverse tipologie di blocchi all'interno di un testo

- ✘ **Paragrafi:** `<P>Paragrafo</P>`
Il testo viene **formattato** automaticamente
- ✘ **Pre:** `<PRE>Blocco di testo</PRE>`
Il testo **NON viene formattato** automaticamente
- ✘ **Sorgente:** `<XMP>Testo sorgente</XMP>`
Il testo viene **reso senza essere interpretato**

26

Formattazione automatica



- I TAG di formattazione automatica supportano l'attributo **align** che può assumere quattro possibili valori:
left center right justify
- Esistono TAG di formattazione forzata:
 -
 a capo su nuova linea
 - <hr> disegna una linea orizzontale

27

Formattazione automatica



- Esistono 6 livelli di importanza per scrivere titoli:
 - <H1>Titolo di primo livello</H1>
 - <H2>Titolo di secondo livello</H2>
 - <H3>Titolo di terzo livello</H3>
 - <H4>Titolo di quarto livello</H4>
 - <H5>Titolo di quinto livello</H5>
 - <H6>Titolo di sesto livello</H6>

28



- Esistono 2 principali tipi di liste:
 - liste unordered
 - liste ordered
- Parti di una lista:
 - <LH> </LH> list header
 - list item
- Attributi del tag :
 - `type` valori possibili:
disc circle square i A a 1
 - `value` valore possibile: numero iniziale



- Il testo può essere formattato usando i tag:
 - **Bold grassetto**
 - <I> *Italico corsivo*
 - <TT> Teletype monodimensionale
 - <BLINK> **Testo intermittente**
 - **Testo da enfatizzare**
 - <CITE> *Citazione*

Collegamenti Iper testuali



- Un collegamento (link) è costituito da
 - ✗ una sorgente, es. scritta su cui cliccare
 - ✗ e una destinazione, es. oggetto visualizzato come risultato del clic
- La direttiva HTML usato nella definizione dei collegamenti prende il nome di:
`<A>` anchor (ancora)

31

La direttiva <A>



- La sintassi per per la direttiva anchor è:
`sorgente`
- **sorgente** è una parte di documento
- **destinazione** può essere:
 - un documento esterno
 - un riferimento all'interno dello stesso documento
 - un riferimento all'interno di un documento esterno

32

Riferimenti Assoluti/Relativi



- I riferimenti ai documenti esterni possono essere:

Assoluti: si indica il completo URL

es. `http://www.cs.unibo.it/~zavattar`

Relativi: si indica il path (cammino) ed il nome del file a partire dalla cartella attuale

es. `/SottoCartella/paginaDiProva.html`

33

Riferimenti interni a documenti



- I riferimenti all'interno di documenti sono identificati con nomi unici
- I nomi dei riferimenti vengono definiti tramite uso dell'attributo `name`

```
<A name="titolo"><H1>Capitolo</H1></A>
```

- Questi nomi possono essere usati nella specifica delle destinazioni di link preceduto dal simbolo `#`

```
<A href="#titolo">Vai al capitolo</A>
```

```
<A href=ExtDoc#titolo>Rif. esterno</A>
```

34