



Il World Wide Web



Sommario

- Il World Wide Web
- HTML
 - Cos'è
- Pubblicazione pagine web





World Wide Web



Il World Wide Web

- Nel 1991 Tim Berner Lee sviluppa al CERN un sistema per la condivisione di informazioni utili per la ricerca che prende il nome di World Wide Web (o web o WWW)
 - L'uso del web si estende dal mondo scientifico a quello dell'educazione a quello delle organizzazioni governative e commerciali
 - Nel 1993 viene realizzato il primo prototipo di browser MOSAIC
-



Il World Wide Web

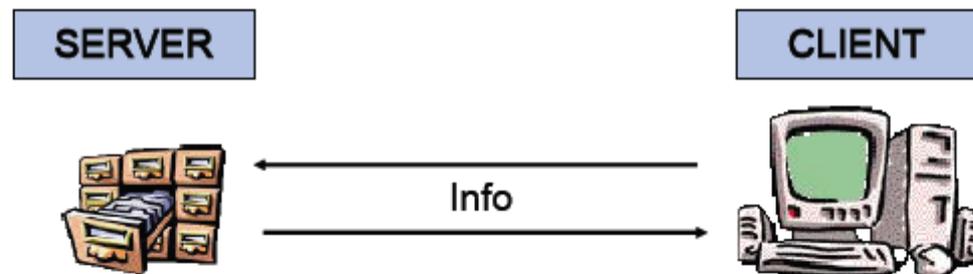
- Rappresenta la “ragnatela” di documenti collegati tra loro attraverso la rete
- Le informazioni sono organizzate in siti
 - un sito è un contenitore logico nel quale l’autore ha inserito contenuti di argomento coerente gli uni con gli altri
- I documenti sono disponibili attraverso i server (**server Web** o **http server**)
- L’applicazione che permette di effettuare le richieste di accesso ai documenti è il **browser Web** (che è un client)
- Le pagine web sono scritte in un particolare linguaggio: **HTML**
- Il protocollo utilizzato per la comunicazione è **l’HTTP**



Protocollo HTTP

Il protocollo HTTP

L'HTTP (HyperText Transfer Protocol)



- Protocollo di tipo applicativo (livello 5)
 - ❑ costruito al di sopra di TCP-IP
 - ❑ usa un approccio Client-Server
 - ❑ è alla base del funzionamento del WWW
 - ❑ ogni interazione prevede esclusivamente una richiesta e una risposta



Il WWW: funzionamento

- L'utente, attraverso il browser (il client), specifica un **indirizzo** (logico o numerico) della pagina che vuole consultare
- Il browser invia la richiesta al server che gliela manda (o spedisce un messaggio di errore)
- Il browser visualizza la pagina all'utente
- La pagina può contenere, oltre al testo anche immagini, suoni, animazioni, filmati etc. Inoltre la pagina contiene punti di collegamento (i **link**) ad altre pagine o parti interattive.

Nota: la comunicazione avviene seguendo le regole del protocollo HTTP



WWW: il lato server

- Un insieme di documenti che trattano uno stesso argomento forma un **sito Web**
- Le pagine con i relativi contenuti sono organizzate in file memorizzati sul computer del server HTTP
 - – es: <http://www.dima.unige.it/~pascarel/index.html>
- Il server HTTP si occupa di accettare le richieste da parte dei client HTTP e di fornire loro i documenti richiesti, se esistono, o dei messaggi di errore



WWW: il lato client, i browser

- E' il programma client che ci consente di richiedere una particolare risorsa (ad es. una pagina web) ad un server utilizzando il protocollo HTTP
 - il browser riceve i risultati, interpreta il linguaggio HTML e produce la pagina a video così come siamo abituati a vederla
- Mozilla, Explorer, ...

- Nelle attuali versioni dei browser convergono tutti gli applicativi per la comunicazione
 - client FTP
 - client telnet



Gli indirizzi del web

- **URL** (Uniform Resource Locator) è uno standard per il formato degli indirizzi delle risorse sul web
 - individua in modo univoco l'host
 - utilizza l'associazione di nomi logici agli indirizzi numerici : DNS
- L'accesso alle pagine del www avviene digitando sul browser un indirizzo specificato da tre parti
 - il protocollo che vogliamo utilizzare per la connessione che identifica il tipo di server a cui ci si vuole connettere
 - ◆ http, ftp, telnet, mailto
 - indirizzo del server
 - ◆ nome logico o indirizzo IP
 - ◆ spesso gli indirizzi dei server web iniziano con www
 - la risorsa che si richiede



Gli indirizzi del web - esempio

- es. <http://www.dima.unige.it/index.html>

indica il protocollo che
vogliamo utilizzare per la
connessione

indirizzo del server

la risorsa che si richiede



Pagine web

- **Tipi di file contenuti in un sito:**
 - file html, file php (per la generazione dinamica di pagine web)
 - immagini, suoni, filmati, file di testo

 - **Statiche**
 - sempre le stesse informazioni (solo HTML)

 - **Dinamiche**
 - generate dal server web su richiesta. Le informazioni visualizzate possono cambiare

 - **Attive**
 - non del tutto predefinite, inviate dal server web e modificate localmente. Dipendono da una serie di fattori, fra cui l'interazione applicazione-utente (con Java e Javascript)
-

Fattori che influenzano la velocità



- Dimensione del file in cui il contenuto della pagina è memorizzato
 - La velocità della propria connessione con il provider
 - La velocità della connessione che il provider riesce ad instaurare con il server su cui risiedono effettivamente i files
 - Il carico a cui è soggetto in quel momento il sito interessato

 - Esistono alcune strategie per velocizzare il trasferimento di una pagina web dal server web cui si è connessi.
-



Cache, proxy server

- I browser hanno una memoria **cache** nella quale sono caricate le ultime pagine visitate dall'utente
 - prima di richiedere al server la pagina web il browser verifica se è contenuta nella sua cache

- I **proxy server** sono nodi speciali all'interno della rete che possono svolgere il ruolo di ponti nella connessione in rete
 - il client si connette al proxy (server per il client)
 - il proxy richiede al server web l'informazione che poi passa al client
 - il proxy svolge un ruolo di cache per tutti i client che lo utilizzano
 - ◆ sono cioè memorizzate in una cache di grandi dimensioni le pagine usate recentemente dagli utenti che passano attraverso i proxy



Cookies

- HTTP non prevede un concetto di sessione
 - ogni elemento che deve essere caricato corrisponde ad una nuova connessione TCP verso il server
 - il server non ha alcun elemento per comprendere quali richieste arrivino da quale utente

 - Il **cookie** è un file di testo scritto dal server web sul calcolatore del client
 - quando il client si connette al server web invia il cookie: tale meccanismo consente al server di riconoscerci

 - Facilmente eliminabili mediante il browser
-



Novità introdotte dal web

- Ipertestualità e archivio
 - Multimedialità
 - www.repubblica.it , www.corriere.it
 - Interattività
 - il lettore parla alla redazione
 - Aggiornamento continuo

 - Come sito di riferimento basta pensare a www.repubblica.it
 - per alcuni la storia del giornalismo online inizia con la nascita di Repubblica. it
 - ◆ il quotidiano mette in rete la sua versione online nel gennaio del '97
-



Iper testo

- Per ipertesto s'intende un testo contenente termini di collegamento o **link** che rimandano ad altri argomenti correlati che possono essere presenti in una pagina sullo stesso sito o un altro sito
 - i link consentono di spostarsi all'interno dei documenti consentendo un'organizzazione completamente innovativa dell'informazione
 - i link possono contenere collegamenti ad
 - ◆ altre pagine
 - ◆ immagini, suoni
 - ◆ programmi eseguibili



Logica ipertestuale

- Logica sequenziale vs logica ipertestuale

 - Un testo tradizionale è basato su una logica di tipo sequenziale
 - chi legge un testo ripercorre insieme all'autore un percorso già impostato adeguandosi al suo ordine logico e cronologico

 - La caratteristica fondamentale di un ipertesto è di non prevedere un unico percorso di accesso alle informazioni ma di adeguarsi alle scelte, alle conoscenze ed alla logica dei fruitori
 - logica più simile al nostro modo di ragionare
-



I link nel giornalismo

“I link nella scrittura online rappresentano dunque un elemento dello stile e una scelta editoriale di primaria importanza. Tanti link a pioggia sono l'equivalente di un pezzo confuso e raffazzonato sul giornale. Pochi link selezionati che contribuiscono alla comprensione e all'approfondimento del pezzo sono il sale del giornalismo online” (Marco Pratellesi)



Multimedialità

- La **multimedialità** è la compresenza e interazione di più mezzi di comunicazione in uno stesso supporto o contesto informativo
 - per comunicare un'informazione riguardo a qualcosa ci si avvale di molti *media*: video, fotografie, immagini, testo, audio
 - ad es. sul sito di repubblica c'è non solo una sezione dedicata alle foto all'audio e al video ma anche un continuo rimando tra queste
 - ◆ a partire dalla homepage il lettore può decidere di leggere vedere o ascoltare la notizia direttamente con un clic



Aspetti tecnologici

- Per ogni tipo di contenuto multimediale vi sono alcuni passaggi obbligati che richiedono un'opportuno supporto tecnologico
 - **authoring**
 - ◆ programmi che permettono di creare oggetti multimediali senza doversi occupare dei dettagli tecnologici
 - **memorizzazione**
 - ◆ cd, dvd
 - **trasmissione e distribuzione**
 - **rappresentazione**
 - ◆ Prodotti software capaci di interpretare i formati
 - **protezione**
 - ◆ tecniche di protezione che dovrebbero limitare la fruizione agli aventi diritto
 - ◆ watermarking
-

Plug-in

- Specialmente per quanto riguarda i dati multimediali la gestione dei formati è in continua evoluzione ovvero si diffondono nuovi formati o si richiede un loro aggiornamento
 - un browser non si limita a visualizzare le pagine web ma e' diventato un applicativo per la riproduzione dei vari media digitali
 - il browser si avvale dei plug-in
- Plug-in: componenti software che si “inseriscono” all’interno del browser aumentandone le potenzialità consentendogli di svolgere nuove funzioni o integrare nuovi media
 - i plugin più diffusi riguardano: QuickTime, Flash, Real Player, PDF





Ipermedia

- Con l'inserimento di contenuti multimediali all'interno di un ipertesto, si è cominciato a utilizzare il termine ipermedia, per evidenziare appunto che il sistema non è composto di solo testo
 - **ipertesto**: arricchire un testo con riferimenti interni ed esterni verso altri testi
 - **multimedialità**: arricchire un testo usando altri canali di comunicazione
 - ipertesto multimediale: **IPERMEDIA**



Pericoli introdotti dal web

- Cercare informazioni nel web diventa un'abitudine per il giornalista
 - Attenzione a dove e quando si trovano le informazioni, che possono poi risultare utili in un momento successivo
 - importante organizzare i siti nei preferiti (bookmarks)
 - Troppa informazione: come scegliere?
 - ogni ricerca va effettuata con un numero sempre maggiore di termini
 - Troppa informazione inaffidabile: come verificarla?
 - cfr. “come copiare da Internet” di Umberto Eco
-



Ricerca di Informazioni in Rete

- Il web comprende moltissime sorgenti di informazione a volte male organizzate, ridondanti, frammentarie, non aggiornate,...
- Per questo motivo esistono servizi (ovvero siti dedicati) che permettono di navigare in modo ragionato:
 - portali web
 - motori di Ricerca
 - newsgroup
 - directory e cataloghi



Portali web

- Contengono diverse sezioni in cui è possibile accedere ad informazioni sui diversi argomenti trattati all'interno del portale
- I portali generalisti contengono diversi tipi di informazione
 - ultime notizie, sport, cinema, ...
 - Yahoo!, Virgilio
- I portali tematici sono siti che contengono informazioni relative ad uno specifico argomento
 - www.univr.it



Motori di Ricerca

- Un motore di ricerca è un servizio di rete che ha la funzione di indicizzare automaticamente le pagine del web
- Consentono ad un utente di ricercare tutte le risorse che contengono una particolare frase o parola
 - si accede ad un motore di ricerca come ad ogni altra pagina tramite il corrispondente indirizzo URL
 - il servizio solitamente e' gratuito
 - ◆ i gestori del sito finanziano la loro attività tramite la pubblicità
- Il numero di risorse che i motori di ricerca conoscono cresce nel corso del tempo
 - gli "spider" indicizzano continuamente i contenuti dei siti web





Newsgroup

- Idea di condividere le proprie conoscenze ed esperienze consentendo a chi lo desidera di esprimere la propria opinione o di porre quesiti
 - Esistono newsgroup su diversi argomenti e chi vuole può aprirne di nuovi
 - gli argomenti sono tra i più disparati
 - Chiunque può inviare un messaggio o leggere e rispondere a messaggi altrui
 - Molti siti offrono la possibilità di accedere ai gruppi di discussione attraverso le pagine web
-



Directory e Cataloghi

- Esistono siti che oltre al servizio di ricerca forniscono un vero catalogo (di parte) delle informazioni disponibili sul web
 - Il catalogo raccoglie i link ad altri siti utilizzando diversi criteri
 - suddivisione per argomenti e sotto-argomenti
 - ordine alfabetico
 - Partendo da una categoria ad es. “Art” è possibile trovare tutti i siti censiti che parlano del gruppo “the Eagle” selezionando le sotto-categorie “Music”, “E”,...
 - si ottiene un numero nettamente inferiori di siti che parlano dell’argomento rispetto ad una ricerca effettuata con Google
 - es. Yahoo, Altavista, Virgilio
-



Creazione di un sito Web

- Avviene in tre fasi:
 - strutturazione e composizione delle pagine che lo costituiscono
 - pubblicazione del sito nel Web
 - presentazione del sito ai motori di ricerca
-



Pubblicazione del sito Web

- Per pubblicare un sito occorre “inserirlo” presso un server sempre accessibile dalla rete. Il server deve essere dotato di un proprio nome e dominio (indirizzo URL) a cui far riferimento per individuare le nostre pagine
 - se siamo amministratori di un server connesso e dotato di un suo nome di dominio allora inseriamo le pagine del sito in una posizione opportuna sui dischi del server
 - ◆ FTP
 - se apparteniamo ad un’istituzione che dispone di un suo sito dovremo rivolgerci all’amministratore
- E’ possibile usare lo spazio Web che i provider mettono a disposizione gratuitamente oppure acquistare lo spazio necessario (sempre da un provider)



Segnalazione ai motori di ricerca

- Il metodo più efficace consiste nel far ‘indicizzare’ il nostro sito da un motore di ricerca
 - l’operazione di indicizzazione avviene automaticamente attraverso opportuni software detti **spider** (raggi) che visitano continuamente il Web saltando da una pagina all’altra attraverso i link e classificandone il contenuto
 - richiesta esplicita (a pagamento)
- Per essere visibili è importante essere ‘linkati’ da altri siti
- E’ fondamentale aiutare l’indicizzazione definendo opportunamente le impostazioni dei tag <META> nel codice html della pagina
 - es.
`<META NAME="keywords" CONTENT ="università">`



Strutturazione di un sito Web: HTML

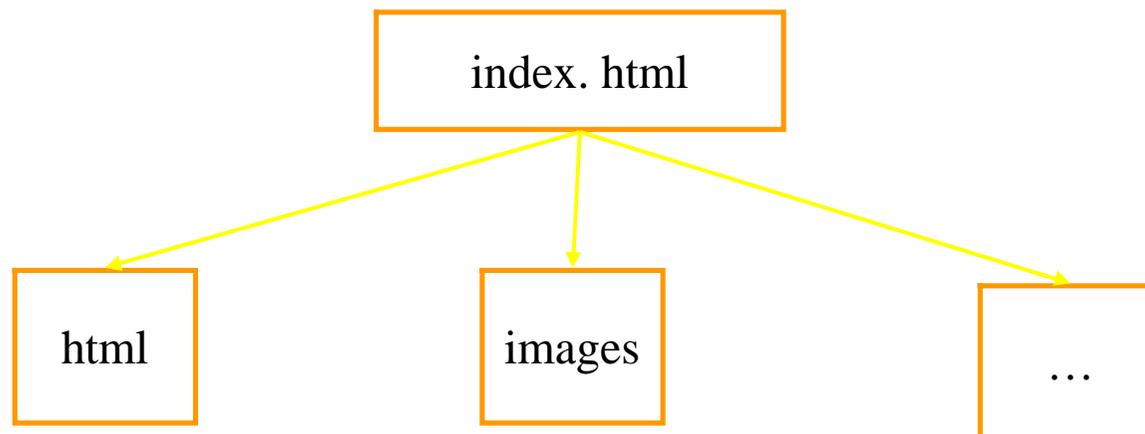


Strutturazione di un sito Web

- Pianificare il contenuto di ciascuna pagina
- Definire l'aspetto grafico
- Definire i riferimenti ad altre pagine
- Definire gli strumenti di sviluppo:
 - HTML
 - XML
 - Linguaggi di scripting (Java, Javascript, CGI)
- Un sito viene creato con programmi di authoring tipo DreamWaver o GoLive! che facilitano lo svolgimento di queste fasi



Schema di un sito Internet



Un sito internet mantiene una struttura gerarchica. Questo perché dobbiamo cercare di mantenere raccolti i files che sono tra di loro omogenei, quindi tutti i files html nella cartella “html”, tutte le immagini dentro alla cartella “images”. Per mettere, ad esempio, a disposizione un font o un documento, possiamo creare una cartella dedicata, chiamandola, per esempio, “download”



Il linguaggio HTML

- HTML = Hypertext Markup Language
- Linguaggio per la costruzione delle pagine web
 - definisce come devono essere visualizzate le pagine
 - permette di realizzare legami ipertestuali tra pagine diverse
- Linguaggio di markup
 - indica come disporre gli elementi all'interno di una pagina mediante delle istruzioni dette "tag"
 - i browser leggono il file html, interpretano i tag e visualizzano la pagina con le giuste formattazioni

Nota: una pagina HTML altro non è che un normalissimo file di testo, che fa riferimento ad altri elementi separati, come immagini, filmati, file multimediali.



I tag

- Le istruzioni HTML sono formate da parole chiave racchiuse tra i simboli “<” e “>”
- Si introducono dei tag che hanno il formato
 - **<comando> </comando>**
- Tutto quello racchiuso tra le coppie di tag deve essere interpretato secondo quanto specificato dal comando
- Esistono tag che non vogliono la chiusura
 - **
**
 - **<hr>**



Struttura di una pagina HTML

- Template di una pagina web

```
<html>
```

```
  <head>
```

```
    <title> titolo </title>
```

```
  </head>
```

```
  <body>
```

```
    contenuto del documento
```

```
  </body>
```

```
</html>
```



Struttura di una pagina HTML

- **<html> </html>**: contenitore del documento
 - è chiuso a fine documento
 - contiene due sezioni
 - ◆ `<head> </head>`
 - ◆ `<body> </body>`
 - **<head> </head>**: contiene informazioni quali
 - titolo del documento
 - fogli di stile
 - scripting
 - **<body> </body>**: racchiude il contenuto
-



Il tag head

- Il tag `<head>` contiene informazioni non visualizzate nel corpo della pagina
 - i motori di ricerca classificano i siti
 - queste informazioni servono ai motori di ricerca
- Il tag `<title>` racchiude il titolo della pagina
 - visibile nella barra del titolo
 - il testo contenuto può comparire anche in seguito alla ricerca in un motore, come titolo del link. Sarà dunque importante impostarlo in modo pertinente; es:

```
<head>
```

```
    <title> Il sito di Anna <title>
```

```
</head>
```



Il tag meta

- Il tag `<meta>` contiene informazioni non visualizzate nel corpo della pagina

- si tratta di una serie di parole chiave e descrizioni, che aiutano i motori di ricerca a classificare il sito; la sintassi è

```
<meta name="..." content = "...">
```

- Ad es. il meta-tag "**description**" permette di impostare una descrizione sintetica del sito stesso
 - ♦ anche in questo caso, la descrizione compare talvolta nei risultati della ricerca

```
<meta name = "description" content = "home page di Annalisa Pascarella">
```

- il meta-tag "**keywords**" permette di indicare alcuni contenuti relativi al sito stesso



Elenco dei meta-tag

Name	Content
description	descrizione sintetica del sito
keywords	descrizione del sito attraverso parole-chiave
author	autore della pagina
copyright	ribadisce diritto d'autore sul contenuto della pagina
generator	editor html utilizzato
lingua	utilizzata all'interno language della pagina



Il Doc type

- DTD = document type definition
 - prima di <html> fornisce le specifiche del documento. Non è obbligatorio
 - di solito generato automaticamente dall'editor. In mancanza, il documento è identificato come HTML 2.0
- es:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
http://www.w3.org/TR/html4/loose.dtd >
```

 - ♦ HTML: il tipo di linguaggio utilizzato è l'HTML
 - ♦ PUBLIC: il documento è pubblico
 - ♦ W3C: il documento fa riferimento alle specifiche del W3C,
 - ♦ - : le specifiche non sono registrate all'ISO. Se lo fossero, sarebbe "+",
 - ♦ DTD HTML 4.01 Transitional: il documento fa riferimento a una DTD, a seguire c'è l'indirizzo del DTD;
 - ♦ EN: la lingua con cui è scritta la DTD è l'inglese



Un pò di utili tag

- `
`
 - per andare a capo
 - non vuole il tag di chiusura!
- `<hr>`
 - per tracciare una linea orizzontale
- `` `` (`<i>` `</i>`)
 - per visualizzare in grassetto (in corsivo)
- `<!-- questo è un commento -- >`



Attributi

- Molti tag hanno degli attributi. La sintassi corretta per un tag è' infatti

```
<tag attributo1="val" attributo2="val"> testo <tag>
```

- es.

```
<body bgcolor="blue" background="img.jpg"> testo  
</body>
```

- Si tende ad avere il contenuto separato dallo stile
 - i fogli di stile (CSS) contengono le regole di formattazione per i vari tag



Tag di formattazione

- `<p> </p>`
 - delimita un paragrafo
 - lascio spazio sopra e sotto
 - `<div> </div>`
 - non lascia spazio prima e dopo la chiusura
 - va a capo
 - ` `
 - elemento inline, non va a capo
-



Elenchi

- ``
 - tag da utilizzare per aprire un elenco ordinato (ol = ordered list)
 - gli elementi sono individuati dal tag ``

 - ``
 - tag da utilizzare per aprire un elenco non ordinato (ul = unordered list)
 - gli elementi sono individuati dal tag ``
-



Elenchi – esempio

- `<ol type = "a" >`
 - ` primo elemento`
 - ` secondo elemento```
 - l'attributo `type` può assumere i valori "1", "i", "I", "a", "A"
 - ◆ è un attributo *case-sensitive*

 - `<ul type="disc" >`
 - `primo elemento`
 - `secondo elemento```
 - l'attributo `type` definisce il tipo di bullet e può assumere i seguenti valori "disc", "circle", "square"
-



Immagini

- ``
 - ❑ tag per inserire immagini
 - ❑ i browser supportano file .jpg, .gif, .png
 - ❑ il valore dell'attributo `src` è il pathname del file
 - ◆ il pathname del file può essere assoluto o relativo
 - ❑ esistono diversi attributi per definire le proprietà delle immagini
 - ◆ dimensioni
 - ◆ posizione rispetto al testo
 - ◆ dimensione bordo dell'immagine
 - ❑ non vuole chiusura



Immagini - esempio

```

```

dimensione
immagine

allineamento



Link

- I link sono "il ponte" che consente di passare da un testo all'altro: all'interno della pagina web i link ci consentono di muoverci da un documento all'altro del web
 - si parla di **ipertesti**
- I link sono definiti con il tag `<a>`
 - ` W3C `
 - ♦ il parametro href del tag permette di specificare l'URL cui il legame deve portare
 - ♦ il testo tra `<a>` e `` e' usato come label per il link
 - ♦ è indifferente che la destinazione sia una pagina HTML di un sito, un'immagine, un file pdf , un file zip, o un file exe: il meccanismo del link funziona allo stesso modo indipendentemente dal tipo di risorsa; poi il browser si comporterà in modo differente a seconda della risorsa



Link - esempio

```
<a title="vai ai link"  
  href="/sito/link.html"> Link </a>
```

etichetta

realizza il collegamento
alla pagina link.html

```
<a href="mailto:pascarel@dima.unige.it">  
  Contatti  
</a>
```

si aprirà il client
di posta predefinito



Pathname

- Se il file si trova nella stessa directory
 - basta specificare il nome del file
- Se il file si trova in una cartella di livello inferiore alla posizione corrente
 - ./cartella/nomefile.html
 - per tornare su di un livello “../nomefile.html ”
- Il path "/" si riferisce alla root della directory, ossia all'home page.
- Se il link è relativo ad una pagina contenuta in un differente sito Web si deve ricorrere ai path assoluti, ossia si deve dare l'intero indirizzo della pagina web (URL)
 - `W3C`



Link interni alla pagina

- E' possibile definire link verso documenti interni alla stessa pagina
 - Per definire dei link interni bisogna compiere due operazioni
 - definire il target (cioè il punto di arrivo del link):
` questo è il target`
 - definire il link con il solito href secondo questa modalità:
`link verso un target interno`
 - Bisogna ricordarsi di utilizzare il simbolo # nell'href verso un target, inoltre i target possono essere sfruttati anche se usiamo un href da un documento esterno
-



Riferimenti

■ Dispense docente

- <http://www.dima.unige.it/~pascarel/html/giornalisti.html>

■ Siti interessanti

- ipertesti
 - ♦ <http://www.mediamente.rai.it/HOME/BIBLIOTE/intervis/1/landow.htm>
- “come copiare da Internet” (Umberto Eco)
 - ♦ <http://espresso.repubblica.it/dettaglio-archivio/1252511>
- HTML
 - ♦ www.w3.org/MarkUp/Guide/
 - ♦ www.html.it
 - ♦ <http://www.w3schools.com/html/default.asp>