

Tecnologie di Sviluppo per il Web

Introduzione Risorse e URI

versione 2.1

Questo lavoro è concesso in uso secondo i termini di una licenza Creative Commons
(vedi ultima pagina)

G. Mecca – mecca@unibas.it – Università della Basilicata



Risorse e URI >> Sommario

Sommario

- Premessa
- Il Linguaggio HTML - Cenni
- Tipi MIME
- URI
- Riferimento ad URI
- Il Protocollo file
- Configurazione del Server HTTP

G. Mecca - Tecnologie di Sviluppo per il Web

2



Premessa

- Il World Wide Web
 - ⇒ piattaforma applicativa di estensione mondiale
- Inizialmente
 - ⇒ sviluppo “disordinato” delle tecnologie
- Ma, molto presto...
 - ⇒ ci si è accorti che era necessario imporre standard unici a livello mondiale



Premessa

- L'attività di standardizzazione
 - ⇒ condotta prevalentemente da due organizzazioni
- World Wide Web Consortium – W3C
 - ⇒ consorzio del WWW, si occupa di standardizzare le tecnologie collegate ai contenuti
- Internet Engineering Task Force – IETF
 - ⇒ si occupa di standardizzare le tecnologie collegate alla rete

Risorse e URI >> Premessa 

Premessa

- Standard di riferimento
 - ⇒ linguaggio HTML
 - ⇒ tipi MIME
 - ⇒ URI
 - ⇒ protocollo HTTP

G. Mecca - Tecnologie di Sviluppo per il Web   5

Risorse e URI >> Il Linguaggio HTML - Cenni 

Il Linguaggio HTML - Cenni

- HTML: HyperText Markup Language
 - ⇒ è il linguaggio con cui vengono costruite la maggior parte delle risorse Web
- Ne diamo solo dei cenni (>>)
- Idea
 - ⇒ costruire pagine (visualizzare testo)
 - ⇒ di carattere ipertestuale (collegamenti)
 - ⇒ e multimediale (immagini e altri media)

G. Mecca - Tecnologie di Sviluppo per il Web   6



Il Linguaggio HTML - Cenni

○ Obiettivo finale

- ⇒ descrivere al client come visualizzare e come gestire l'interazione dell'utente con la pagina
- ⇒ es: selezione dei collegamenti

○ Marcatore

- ⇒ per dare indicazioni al browser, vengono utilizzati marcatori ("tag")
- ⇒ ogni marcatore ha una opportuna sintassi ed una semantica



Il Linguaggio HTML - Cenni

```

<HTML>
<HEAD>
  <TITLE>Programma del Corso di Sviluppo Web</TITLE>
</HEAD>
<BODY>
  <H1>Corso di Tecnologie di Sviluppo per il Web</H1>
  <P>Questa &grave; una pagina di <B>esempio</B>.</P>
  <P>
    <A href="altraPagina.html">
      Questo e' un collegamento ad un'altra pagina
    </A>
  </P><HR>
  <I>G. Mecca, ultima modifica Marzo 2003
</BODY>
</HTML>
  
```

riferimento ad immagine

collegamento ad altra pagina
(riferimento ed ancora) >> /tmp/esempio.html



Il Linguaggio HTML - Cenni

○ Riassumendo

- ⇒ una pagina Web è una collezione di risorse
- ⇒ composta di una risorsa principale, che contiene il codice HTML (es: esempio.html)
- ⇒ più una serie di altre risorse “interne”, come immagini ed altri oggetti multimediali (es: lavagnaWeb.gif)
- ⇒ ad ogni pagina sono normalmente collegate altre pagine



Tipi MIME (“MIME Types”)

- Non tutte le risorse sono risorse HTML
- Tipi MIME
 - ⇒ Multipurpose Internet Mail Extensions
 - ⇒ descrivono il contenuto delle risorse
 - ⇒ diversi formati (es: testo, HTML, gif, jpeg...)
 - ⇒ identificatori standard per vari tipi di file
- Standard IETF (RFC 2045, RFC 2046)
 - ⇒ mantenuti da IANA



Tipi di Contenuto ("Content-Types")

- text
 - ⇒ text/plain
 - ⇒ text/html
 - ⇒ text/xml ...
- image
 - ⇒ image/gif
 - ⇒ image/jpeg
 - ⇒ image/png ...
- application
 - ⇒ application/pdf
- ⇒ application/zip
- ⇒ application/msword
- ⇒ application/X-...
- video
 - ⇒ video/mpeg ...
- audio
- multipart
 - ⇒ multipart/form-data ...
- message
- model



Tipi MIME ("MIME Types")

- Meccanismo di associazione
 - ⇒ ogni volta che il server invia al client una risorsa specifica nella risposta il tipo MIME
 - ⇒ nel caso di risorse statiche (file) il server sceglie il tipo MIME sulla base dell'estensione del file
 - ⇒ un opportuno file di configurazione viene utilizzato per associare alle varie estensioni di file il tipo MIME corrispondente (>>)



Tipi MIME ("MIME Types")

- Gestione del tipo MIME
 - ⇒ il browser riceve la risposta
 - ⇒ consulta il tipo MIME
 - ⇒ e sulla base del tipo MIME decide come gestire il contenuto della risposta
 - ⇒ es: visualizzazione nella sua finestra o avvio di un plug-in esterno



Uniform Resource Identifiers (URI)

- Sistema di indirizzamento su Web
 - ⇒ stringhe di caratteri ASCII che identificano le risorse disponibili su Web
- Standard IETF (RFC 2396)
- Uniform Resource Identifiers (URI)
 - ⇒ due categorie
 - ⇒ Uniform Resource Locators (URL)
 - ⇒ Uniform Resource Names (URN)



Uniform Resource Identifiers (URI)

- URL

- ⇒ la risorsa è fisicamente accessibile
- ⇒ la stringa descrive il metodo (primario) per accedere alla risorsa

- URN

- ⇒ la stringa non descrive il metodo d'accesso
- ⇒ la risorsa può non essere fisicamente accessibile (es: namespace)

- Ci concentreremo sugli URL



URI

- Forma generale

- ⇒ <protocollo>:<parte-dipendente-dal-protocollo>

- Principali protocolli

- ⇒ http
- ⇒ ftp
- ⇒ mailto
- ⇒ file

Risorse e URI >> URI 

URI

- Esempi:
 - ⇒ <http://www.db.unibas.it/users/mecca/>
 - ⇒ <ftp://lab.unibas.it/pub>
 - ⇒ <mailto:mecca@unibas.it>
 - ⇒ <file:///d:/sites/users/gmecca/index.html>
 - ⇒ <gopher://spinaltap.micro.umn.edu/00/>
 - ⇒ <news:comp.infosystems.www.servers.unix>
 - ⇒ <telnet://melvyl.ucop.edu/>

G. Mecca - Tecnologie di Sviluppo per il Web   17

Risorse e URI >> URI 

Terminologia

- Nucleo del Web (“Web Core”)
 - ⇒ tutte le risorse il cui URI è basato su HTTP o su varianti di HTTP (es: HTTPS)
- Periferia del Web
 - ⇒ risorse il cui URI è basato su protocolli diversi da HTTP (es: FTP, mailto) ma sono direttamente collegate al nucleo
- Ci concentriamo sul nucleo
 - ⇒ URI basati su HTTP

G. Mecca - Tecnologie di Sviluppo per il Web   18



URI basati su HTTP

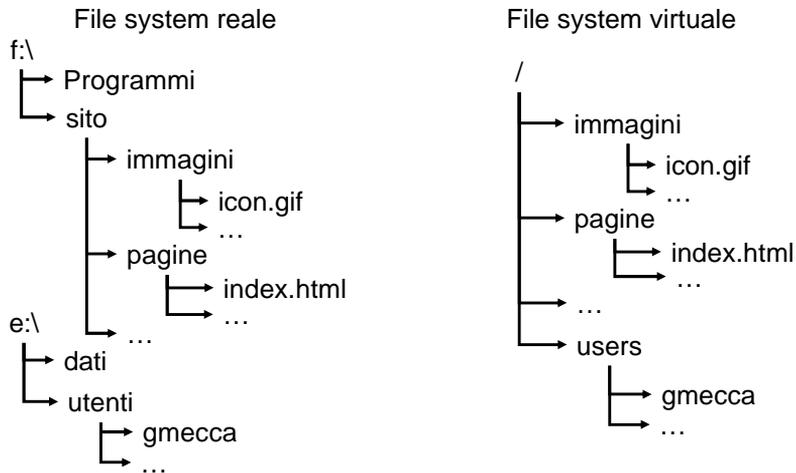
- Contengono riferimenti alle risorse
 - ⇒ es: file su disco
- Organizzazione dei file per il server
 - ⇒ sulla base di un file system “virtuale”
- File system
 - ⇒ regole e moduli per l’accesso ai file in un sistema
 - ⇒ ogni sistema operativo ne ha uno (es: Windows, Linux)



URI basati su HTTP

- File system del server HTTP
 - ⇒ è virtuale perchè non corrisponde all’organizzazione fisica dei file sul disco
 - ⇒ contiene una porzione del disco reale
 - ⇒ riorganizzata “montando” e “smontando” pezzi
 - ⇒ i riferimenti ai file nelle richieste al server HTTP devono essere fatti utilizzando le regole del file system virtuale

URI basati su HTTP



URI basati su HTTP

- File system virtuale del server HTTP
 - ⇒ convenzione di nomi di UNIX (/)
 - ⇒ costruito specificando una radice
 - ⇒ e montando sulla radice ulteriori pezzi (alias)
- Directory radice
 - ⇒ sul file system virtuale si chiama "/"
 - ⇒ bisogna specificare a quale cartella del disco reale corrisponde; es: f:\sito\



URI basati su HTTP

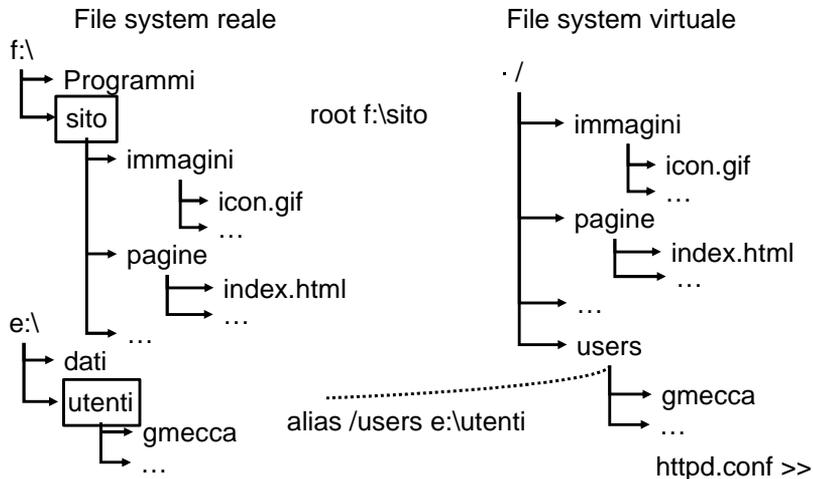
- Una volta montata la radice
 - ⇒ è visibile sul file system virtuale tutto il contenuto della radice
 - ⇒ attenzione ai nomi
- Sottodirectory
 - ⇒ es: f:\sito\immagini\ >> /immagini/
- File
 - ⇒ es: f:\sito\immagini\icon.gif >> /immagini/icon.gif



URI basati su HTTP

- Alias
 - ⇒ percorsi corrispondenti a “punti di montaggio”
- E' possibile montare:
 - ⇒ altre porzioni del file system reale sotto la radice virtuale, specificando dei nomi
es: e:\utenti >> /users/
 - ⇒ da quel momento tutto il contenuto è visibile
es: e:\utenti\gmecca\ >> /users/gmecca/
e:\utenti\index.html >> /users/index.html

URI basati su HTTP



URI basati su HTTP

○ Nota

⇒ il meccanismo degli alias viene utilizzato anche per montare applicazioni oltre che file

○ Idea

⇒ all'applicazione viene associato un URI

⇒ l'URI può non corrispondere ad un file

⇒ richiedendo l'URI viene richiesta l'esecuzione dell'applicazione (non il contenuto del file)

⇒ risposta generata dinamicamente (>>)



URI basati su HTTP

- Forma generale di un URI HTTP
 - ⇒ `http://<server>[:<porta>][/<percorso>][?<query>]`
- Server, porta, percorso
 - ⇒ indirizzo IP del server e porta TCP (80)
 - ⇒ percorso nel file system virtuale
- Query
 - ⇒ lista di coppie “nome=valore” separate da &
 - ⇒ parametri per l'esecuzione di un'appl. (>>)



URI basati su HTTP

- Caratteri riservati nella sintassi
 - ⇒ `; / ? : @ & = + $, < > # % ' ' (spazio)`
`{ } | \ ^ [] ``
 - ⇒ questi caratteri devono essere codificati
- Codifica dell'URI (“URI Encoding”)
 - ⇒ processo di codifica dei caratteri speciali
 - ⇒ formato: `%<hex><hex>` dove `<hex><hex>` è la codifica esadecimale del codice ASCII
 - ⇒ es: spazio, cod ASCII: 32, codifica: `%20`



URI Basati su HTTP

○ Esempi

<http://www.unibas.it/index.html>

<http://www.unibas.it:8080/news/index.html>

<http://www.db.unibas.it/users/mecca/corsi/index.html>

<http://www.db.unibas.it/users/gmecca/icons/dot.gif>

<http://www.aci.it/pagabollo?targa=AD123DE>

<http://www.microsoft.com/redirect.dll?prd=ie&ar=windows>

<http://www.finanze.it/cf.asp?utente=G.%20Mecca>



URI

○ Una particolarità

⇒ riferimenti a “frammenti”: punti definiti in una risorsa

⇒ ormai poco usati

○ Sintassi: #<nome-del-frammento>

⇒ es: <http://www.unibas.it/facolta/index.html#scienze>

○ Semantica

⇒ il browser visualizza la pagina e si posiziona in corrispondenza dell’inizio del frammento



Riferimento ad URI

- Richiesta del client
 - ⇒ è necessario specificare l'URI della risorsa
- Varie categorie di richieste
 - ⇒ richieste esplicite dell'utente; es: l'utente digita l'URI della richiesta nella barra indirizzi
 - ⇒ richieste di collegamenti; es: l'utente seleziona un collegamento ipertestuale
 - ⇒ richieste implicite; es: per visualizzare una pagina HTML è necessaria una immagine



Riferimento ad URI

- Collegamenti e richieste implicite
 - ⇒ URI provenienti dal codice HTML di una pag.

<http://www.unibas.it/didattica/index.html>

<http://www.unibas.it/didattica/facolta/elenco.html>





Riferimento ad URI

- Richieste esplicite
 - ⇒ tipicamente l'utente specifica l'URI completo della risorsa
- Richieste implicite e collegamenti
 - ⇒ nel codice HTML della pagina è possibile specificare un riferimento completo all'URI della destinazione
 - ⇒ oppure un riferimento relativo



Riferimento ad URI

- Riferimento relativo ad un URI
 - ⇒ contiene solo <percorso>[?<query>]
es: facolta/elenco.html
 - ⇒ relativo ad un URI base di riferimento
 - ⇒ come URI base viene utilizzato l'URI della pagina di partenza (es: Didattica)
 - ⇒ il riferimento viene automaticamente espanso dal browser nell'URI completo della destinazione (es: Facoltà)



Riferimento ad URI

○ Esempi:

⇒ nella pagina:

<http://www.unibas.it/didattica/index.html>

⇒ URI reference: “[facolta/elenco.html](#)” >>

<http://www.unibas.it/didattica/facolta/elenco.html>

⇒ URI reference: “[logo.gif](#)” >>

<http://www.unibas.it/didattica/logo.gif>



Riferimento ad URI

○ Espansione dell'URI

⇒ le parti completamente mancanti (protocollo, server, porta) vengono prese dall'URI di base

⇒ il percorso viene completato

⇒ si assume come directory iniziale quella della pagina di partenza

⇒ si aggiunge la porzione di percorso specificato nel riferimento relativo



Riferimento ad URI

○ Altri Esempi:

⇒ nella pagina:

<http://www.unibas.it/didattica/index.html>

⇒ URI reference: “/logo.gif” (nella radice) >>

<http://www.unibas.it/logo.gif>

⇒ URI reference: “../index.html” (.. = padre) >>

<http://www.unibas.it/index.html>



Riferimento ad URI

○ URI reference che termina con una cartella

⇒ e non con un file; es: <http://www.unibas.it/didattica/>

○ Completamento dell'URI

⇒ il server restituisce un file di nome standard

es: [index.html](#) oppure [default.htm](#)

⇒ parametro configurabile

○ Utilizzo tipico: pagina “host”

⇒ pagina principale di un sito; es: <http://www.unibas.it>

⇒ pagina il cui percorso corrisponde a “/”

>> </sviluppoWeb/riferimenti.html>



Terminologia

○ Sito Web

- ⇒ collezione di risorse Web con le seguenti caratteristiche (es: <http://www.db.unibas.it>)
- ⇒ sono collocate allo stesso indirizzo di rete (server e porta dell'URI comuni)
- ⇒ sono raggiungibili tutte attraverso collegamenti ipertestuali a partire dalla pagina principale del sito (percorso "/")

○ Sottositi

- ⇒ porzioni di un sito indipendenti per funzione e gestione (es: www.db.unibas.it/users/mecca)



Il Protocollo File

○ Esiste un protocollo alternativo ad http

- ⇒ il protocollo file://

○ Differenza fondamentale

- ⇒ serve a fare riferimento a file sul file system reale della macchina (non su quello virtuale)

○ Sintatticamente

- ⇒ gli URI si costruiscono in modo simile
- ⇒ è possibile utilizzare la convenzione relativa ai nomi di file del sistema operativo



Il Protocollo File

○ A cosa serve

- ⇒ ad aprire documenti html locali attraverso il browser senza passare per un server http
- ⇒ può essere utile per fare prove veloci del codice html

○ Attenzione alle differenze

- ⇒ le richieste non passano dal server
- ⇒ riferimenti relativi diversi
- ⇒ no espansione automatica



Il Protocollo File

○ Nota

- ⇒ nei sistemi Windows tutti gli URI devono cominciare con / (che rappresenta il “punto di ingresso” nel file system reale)
- ⇒ seguita dal nome del disco e dal percorso
- ⇒ gli slash devono essere comunque diretti
- ⇒ es: `file:///d:/corsi/sviluppoWeb/index.html`

>> `file:///d:/corsi/sviluppoWeb/materialeLatoServer/`



Configurazione del Server HTTP

- Nel corso di queste lezioni
 - ⇒ vedremo progressivamente come installare e configurare un server HTTP
- Prima operazione da effettuare
 - ⇒ installazione dell'applicazione
 - ⇒ es: Internet Information Services dal CD di Windows 2000 Pro/XP Pro
 - ⇒ es: Apache (<http://httpd.apache.org>)



Configurazione del Server HTTP

- Internet Information Services
 - ⇒ pannello di controllo >> installazione applicazioni >> componenti di Windows
 - ⇒ spuntare la voce "Internet Information Services" e proseguire
 - ⇒ si avvierà la procedura di installazione dal CD
 - ⇒ al termine dell'installazione viene creato un nuovo servizio "Pubblicazione su Web"



Configurazione del Server HTTP

○ Apache 2

- ⇒ scaricare la versione installabile da <http://httpd.apache.org/>
- ⇒ installare la versione eseguendo la procedura di installazione ordinaria
- ⇒ al termine dell'installazione viene creato un nuovo servizio "Apache 2"

○ Attenzione

- ⇒ entrambi i servizi ascoltano sulla porta 80



Configurazione del Server HTTP

○ Seconda operazione

- ⇒ configurazione del server

○ Primi parametri da specificare

- ⇒ radice del file system virtuale
- ⇒ eventuali alias corrispondenti ad altre cartelle

○ Internet Information Services

- ⇒ strumento di configurazione grafico

○ Apache

- ⇒ procedura testuale

Configurazione del Server HTTP

- Internet Information Services

- ⇒ Pannello di controllo >> Strumenti di amministrazione >> Internet Information Services

- Apache

- ⇒ file httpd.conf nella cartella conf

- ⇒ normalmente in c:\Programmi\Apache Group\Apache 2

>> IIS, httpd.conf

Configurazione del Server HTTP

- Riassumendo: parametri fondamentali

- ⇒ radice del file system

- ⇒ alias

- ⇒ porta del servizio

- Altri parametri da configurare

- ⇒ documento standard (es: index.html)

- ⇒ tipi MIME



Configurazione del Server HTTP

- Per i file sul disco
 - ⇒ il server HTTP decide il tipo MIME sulla base dell'estensione del file richiesto
- Configurazione >> mime.types
 - ⇒ elenco di estensioni con il relativo tipo MIME
- Per i contenuti dinamici
 - ⇒ nel caso di servizi interattivi, il tipo MIME viene specificato dallo sviluppatore (>>)



Configurazione del Server HTTP

- Configurazione del Logging
 - ⇒ quanti file di log
 - ⇒ formato del file di log (quali informazioni registrare)
 - ⇒ strategia di rotazione

>> httpd.conf – logs
>> IIS - logs



Riassumendo

- Premessa
- Il Linguaggio HTML - Cenni
- Tipi MIME
- URI
- Riferimento ad URI
- Il Protocollo file
- Configurazione del Server HTTP



Termini della Licenza

- This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.
- Questo lavoro viene concesso in uso secondo i termini della licenza "Attribution-ShareAlike" di Creative Commons. Per ottenere una copia della licenza, è possibile visitare <http://creativecommons.org/licenses/by-sa/1.0/> oppure inviare una lettera all'indirizzo Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.