

World Wide Web

Introduzione
 Caratteristiche fondamentali
 URL
 Architettura e funzioni base
 HTML, XML, XHTML
 CSS, XSL
 HTTP

Introduzione

- Il Web e' un sistema ipermediale distribuito
 - sistema ipermediale: permette l'accesso interattivo a collezioni di documenti di diverso tipo (testo, figure, immagini, animazioni, audio, video) collegati tra loro
 - sistema distribuito: documenti conservati su piu' server remoti
- Inventato da Tim Berners-Lee nel 1989 al CERN di Ginevra per promuovere il lavoro collaborativo nel campo della fisica delle particelle
- Prima interfaccia grafica: Mosaic (freeware), 1993, di Marc Andreessen, NCSA (<http://www.ncsa.uiuc.edu/>)
- Ulteriori sviluppi gestiti dal consorzio World Wide Web: W3C (<http://www.w3.org/>)

Caratteristiche fondamentali

- Idea
 - un mondo di informazioni senza confini in cui tutti gli elementi hanno un riferimento per poter essere ritrovati
- Sistema di indirizzamento (URL)
 - per rendere questo mondo possibile nonostante molti protocolli diversi
- Un protocollo richiesta-risposta (HTTP)
 - per trasferire informazioni con l'efficienza necessaria al contesto ipermediale
- Un linguaggio di annotazione (HTML)
 - che tutti i client capiscono, per presentare le informazioni e inserire gli iper-puntatori (hyperlinks)
- Un corpo di dati disponibile su Internet
 - chiunque puo' rendere pubblici i dati mettendoli su un server

URL - Uniform Resource Locator

- Sintassi (RFC 1738): `method://host_domain[:port]/path/filename`
- Struttura dell'URL
 - nome del protocollo (http, ftp, ...)
 - indirizzo DNS (il browser richiede la risoluzione del DNS name)
 - numero di porta (opzionale, default 80)
 - risorsa target (es.nome file)
- Esempi:
 - <http://www.di.uniba.it/~reti/>
 - <http://seldi2.uniba.it:1025/>
 - <ftp://seldi.uniba.it/pub/papers/sew95.ps>
 - <file://lucidi/web.zip>
 - <news:it.lavoro.offerte>
 - <mailto:lanubile@di.uniba.it>

Architettura client/server

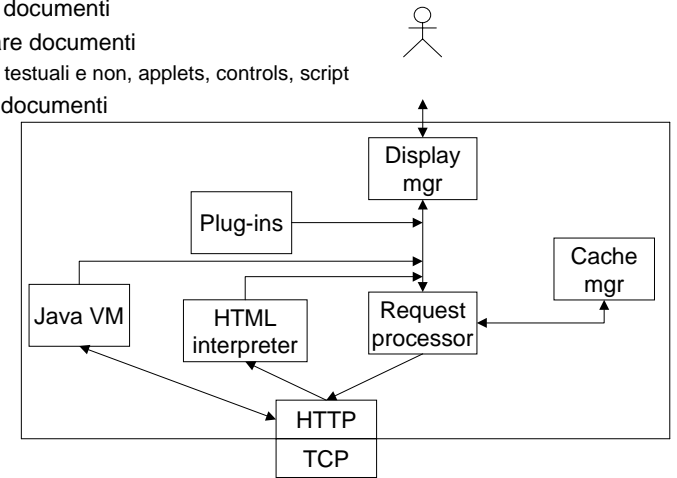
- l'utente seleziona un URL target
- il browser (web client) spedisce una richiesta HTTP al web server
- Il server elabora la richiesta e spedisce la risposta al browser
- Il browser interpreta i comandi HTML e mostra la pagina
- Ogni documento esterno referenziato richiede una richiesta separata



Architettura del browser

Funzioni base

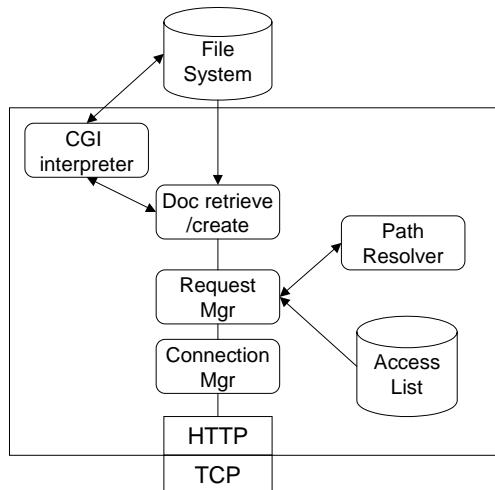
- Ritrovare documenti
- Interpretare documenti
 - entita' testuali e non, applets, controls, script
- Mostrare documenti



Architettura del web server (http daemon)

Funzioni base

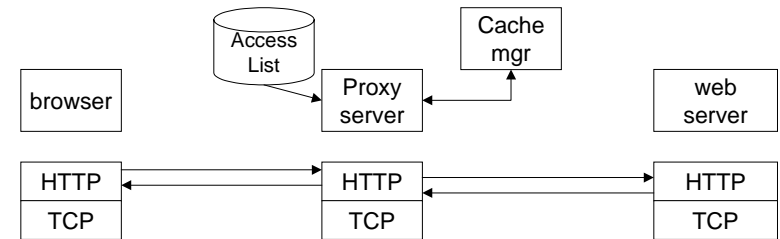
- gestire connessioni con i browser
- gestire richieste e risposte HTTP
- ritrovare/creare documenti
- gestire sicurezza



Proxy server

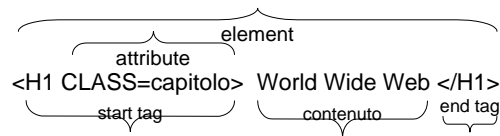
Funzioni base

- caching centralizzata
- firewall: sicurezza mediante filtraggio
- gateway per protocolli non supportati dal browser



Markup Language

- Markup language: linguaggio che contiene comandi espliciti per la formattazione (es. TeX)
- Standard Generalized Markup Language è un metalinguaggio per documenti strutturati (ISO 9979)
- La specifica di un linguaggio specifico in SGML è contenuta in un Document Type Definition (DTD): collezione di elementi e attributi
- Un documento è valido se è conforme al DTD a cui si riferisce



Filippo Lanubile - World Wide Web

9

HTML

- HyperText Markup Language (HTML) è un'applicazione dell'SMGL per ipermedia
- Standardizzazione gestita dal W3C
- Un documento HTML è composto da tre parti
 - una linea (opzionale) contenente informazioni sulla versione HTML
 - una sezione dichiarativa di intestazione (delimitato dall'elemento HEAD)
 - un corpo, che contiene il documento vero e proprio (delimitato dagli elementi BODY o FRAMESET)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<HTML>
<HEAD>
  <TITLE>Il mio primo documento HTML </TITLE>
</HEAD>
<BODY>
  <P>Hello world!
</BODY>
</HTML>
```

Filippo Lanubile - World Wide Web

10

Differenze tra versioni HTML

- HTML 1.0 (1990-1992, TBL)
 - hyperlinks, images, lists
- HTML 2.0 (1994, RFC 1866, IETF & W3C)
 - Active Maps and Images, Forms
- HTML 3.0 (1995)
 - Equations, toolbars, tables
- HTML 3.2 (Jan. 1997, W3C)
 - applets, text flow around images, style sheets
- HTML 4.0 (Dec, 1997, W3C)
 - cascading style sheets, scripts, frames, embedded objects, richer tables, enhanced forms, accessibility for disabled people, internazionalization
- HTML 4.01 (Dec, 1999, W3C)
 - correzioni e piccoli cambiamenti
- XHTML 1.0 (Jan., 2000, W3C)
 - riformulazione dell'HTML 4.0 in XML 1.0

Verificare la lista degli elementi deprecati nelle ultime versioni

Filippo Lanubile - World Wide Web

11

HTML Tag

- I comandi di formattazione sono chiamati TAG e spesso sono in coppia
- Alcuni tags

– <A> 	Anchor (link or name)
– <BODY> </BODY>	Contents
– 	Break
– <FORM> </FORM>	Input form
– <H1> </H1>	Heading level 1
– <HEAD> </HEAD>	Header of a document
– <HR>	Horizontal Rule
– <HTML> </HTML>	The doc type is HTML
– 	List Item
– 	Ordered List
– <P> </P>	Paragraph
– <TITLE> </TITLE>	Document title
– 	Unnumbered list

Filippo Lanubile - World Wide Web

12

Puntatori

- ```
 testo o file grafico
```
- URL assoluta:
    - `<A HREF="http://www.di.uniba.it">Dipartimento di Informatica</A>`
  - URL relativa
    - `<A HREF="public.html">Publications</A>`
  - Anchor points
    - `<A HREF="#background">Background</A>`
  - Possibilità di far riferimento anche a uno specifico frame o istanza di browser dove caricare il documento puntato
    - `<A HREF="http://www.di.uniba.it/" TARGET="maindoc">Dipartimento di Informatica</A>`

## Fogli di stile (stylesheets)

- Obiettivo: separare struttura e presentazione di un documento
- I fogli di stile specificano il formato di presentazione degli elementi di un documento
  - `<STYLE TYPE ="text/css"> H1 { border : thick solid blue } </STYLE>`
- Possono essere di due tipi:
  - Interni  
vengono inseriti direttamente tra i tag `<HEAD>` e `</HEAD>` del documento
  - Esterni (fortemente consigliati)  
lo stylesheet viene invocato all'interno del documento mediante puntatore ad un file esterno, contenente il codice dello stylesheet stesso  
`<link rel="stylesheet" href="style.css" type="text/css">`
- Possono essere applicati in cascata: cascading style sheets (CSS)

## Vantaggi dei fogli di stile

- La possibilità di tenere separato uno stylesheet dai documenti HTML su cui potrebbe essere utilizzato offre svariati benefici :
  - Una volta creato, uno stylesheet esterno può essere riutilizzato più volte
  - Gli autori possono cambiare il "look" ai loro documenti, effettuando singole modifiche solo sui relativi stylesheet esterni.
  - Gli interpreti possono scaricare gli stylesheet selettivamente, basandosi, ad esempio, sulle eventuali descrizioni del mezzo (CSS2, 1998)
 

```
@media speech {
 BODY { voice-family: female }
 H1 { volume: loud}
}
```
- HTML 4.0 depreca l'uso di marcatori di presentazione (es. FONT, B) e impone l'uso dei fogli di stile per formattare i documenti HTML

## XML

- Extensible Markup Language (XML) è un'estensione dell'HTML che permette di definire nuovi elementi
- Creato e standardizzato da W3C (1998)
- XML è stato progettato per descrivere dati con tags non predefiniti a differenza dell'HTML
- Un documento XML è una collezione gerarchica di elementi
 

```
<?xml version="1.0" encoding ="UTF-8"?>
<ARTICLE>
<HEADLINE>Fredrick the Great meets Bach</HEADLINE>
<AUTHOR>Johann Nikolaus Forkel</AUTHOR>
<PARA>
 One evening, just as he was getting his
 <INSTRUMENT>flute</INSTRUMENT> ready and his
 musicians were assembled, an officer brought him a list of
 the strangers who had arrived.
</PARA>
</ARTICLE>
```
- E' necessario utilizzare un XML parser per estrarre la struttura di un documento

## DTD

- Un documento XML sintatticamente corretto si dice *well-formed*
  - Tutti i tags devono essere in coppia
  - Gli elementi devono essere in sequenza o nidificati (children)
- Un documento XML può essere associato a un DTD (Document Type Definition)
 

```
<!DOCTYPE author [
 <!ELEMENT author (name, email)>
 <!ELEMENT name (#PCDATA)>
 <!ELEMENT email (#PCDATA)>
]>
```
- Un documento XML che rispetta la sintassi definita nel suo DTD si dice *valid*

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE author SYSTEM "author.dtd">
<author> ...
```
- XML si sta rapidamente affermando per lo scambio di dati tra sistemi diversi
  - E' necessario utilizzare gli stessi DTD oppure operare trasformazioni tra DTD differenti

Filippo Lanubile - World Wide Web

17

## CSS e XML

- Un documento XML non specifica come deve essere presentato
  - è possibile usare CSS ma esistono altre alternative

```
<?xml-stylesheet href="common.css" type="text/css"?>
<?xml-stylesheet href="modern.css" title="Modern" media="screen"
 type="text/css"?>
<?xml-stylesheet href="classic.css" alternate="yes"
 title="Classic" media="screen, print" type="text/css"?>
<ARTICLE>
<HEADLINE>Fredrick the Great meets Bach</HEADLINE>
<AUTHOR>Johann Nikolaus Forkel</AUTHOR>
<PARA>
 One evening, just as he was getting his
 <INSTRUMENT>flute</INSTRUMENT> ready and his
 musicians were assembled, an officer brought him a list of
 the strangers who had arrived.
</PARA>
</ARTICLE>
```

Filippo Lanubile - World Wide Web

18

## CSS e XML(cont.)

```
/* common.css */
INSTRUMENT { display: inline }
ARTICLE, HEADLINE, AUTHOR, PARA { display: block }

/* modern.css */
ARTICLE { font-family: sans-serif; background: white; color: black }
AUTHOR { margin: 1em; color: red }
HEADLINE { text-align: right; margin-bottom: 2em }
PARA { line-height: 1.5; margin-left: 15% }
INSTRUMENT { color: blue }

/* classic.css */
ARTICLE { font-family: serif; background: white; color: #003 }
AUTHOR { font-size: large; margin: 1em 0 }
HEADLINE { font-size: x-large; margin-bottom: 1em }
PARA { text-indent: 1em; text-align: justify }
INSTRUMENT { font-style: italic }
```

Filippo Lanubile - World Wide Web

19

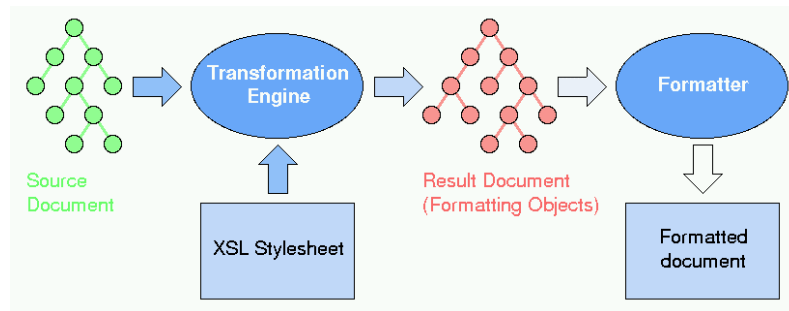
## XSL

- Extensible Stylesheet Language (XSL)
  - è un linguaggio dichiarativo, creato dal W3C e in corso di standardizzazione
  - permette di specificare
    - come trasformare un documento XML
      - trasformazione in un altro documento XML
      - trasformazione in HTML, XHTML, text, o altro linguaggio interpretabile da un dispositivo di presentazione
      - generazione di un indice, ordinamento, selezione
    - come presentare un documento XML
      - XSL fornisce funzionalità più sofisticate rispetto a CSS
  - può essere utilizzato sia in modo server-side che client side

Filippo Lanubile - World Wide Web

20

## Architettura XSL



## Componenti XSL

- XSLT
  - un linguaggio di trasformazione
- XPath
  - un linguaggio per indirizzare parti di documenti XML
- FO
  - un vocabolario di formatting objects con le loro associate proprietà di formattazione
    - Es. formattazione in PostScript o PDF

XSL usa XSLT e XSLT usa XPath

## XSLT

- Una trasformazione XSL
  - accetta per input albero sorgente, rappresentato da un documento XML
  - produce per output un albero risultato, rappresentato generalmente da un altro documento XML (può anche essere HTML o testo)
- Un file XSL è composto da una lista di regole di trasformazione (template rules)
  - una template rule specifica come ogni nodo dell'albero sorgente deve apparire nell'albero risultato
  - Una template rule: pattern + template
  - il pattern (espresso in Xpath) specifica il nodo (radice) dell'albero sorgente a cui applicare la regola di trasformazione
  - il template specifica il nodo (radice) dell'albero risultato da instanziare

## XSLT (cont.)

- XSL Processor
  - Il processore XSLT scandisce l'albero sorgente, partendo dal nodo radice per cercare un matching template rule nel file XLS
  - quando un pattern è individuato nell'albero sorgente, il template corrispondente è generato nell'albero risultato
  - l'istruzione `<xsl:apply-templates/>` permette di applicare le regole di trasformazione in modo ricorsivo ai nodi figli

```
<xsl:template match="para">
 <p><xsl:apply-templates/></p>
</xsl:template>
```

- Input: `<para>Esempio paragrafo</para>`
- Output: `<p>Esempio paragrafo</p>`

## Esempio completo di trasformazione XSL (1)

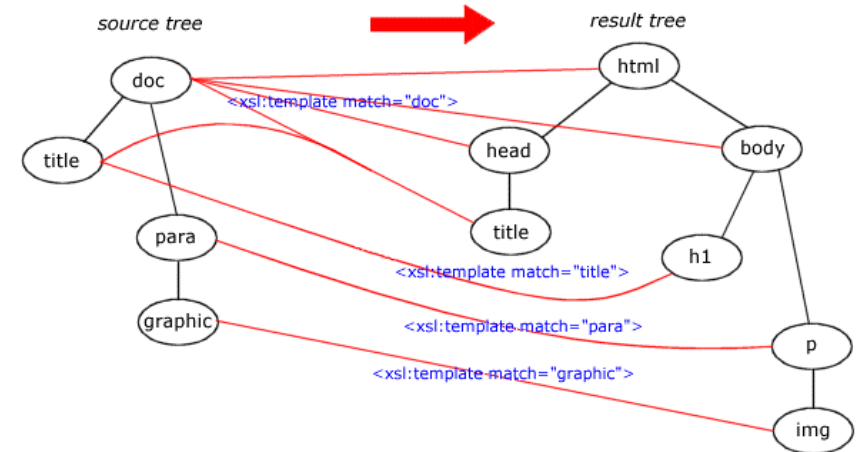
```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
 version="1.0">
 <xsl:output method="html"/>

 <xsl:template match="doc">
 <html>
 <head><title><xsl:value-of select="title"/></head>
 <body><xsl:apply-templates/></body>
 </html>
 </xsl:template>

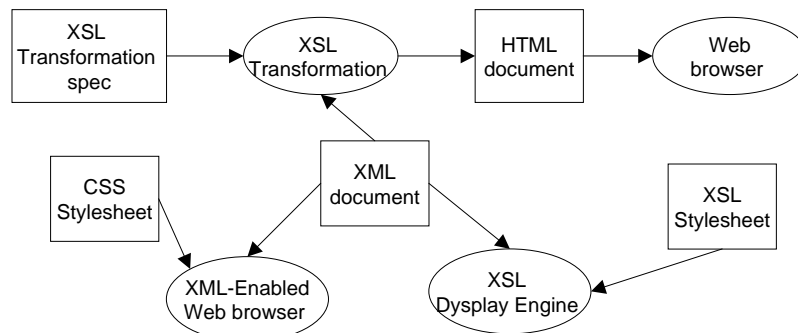
 <xsl:template match="title">
 <h1><xsl:apply-templates/></h1>
 </xsl:template>

 <xsl:template match="para">
 <p><xsl:apply-templates/></p>
 </xsl:template>
</xsl:stylesheet>
```

## Esempio completo di trasformazione XSL (2)



## Opzioni per mostrare documenti XML



## Xlink e XPointer

- XML Linking Language (XLink)
  - Linguaggio per esprimere riferimenti arricchiti a documenti e risorse di rete
  - generalizza HTML A, IMG, LINK
  - possibilità di definire link tra n risorse
  - possibilità di definire più di un comportamento associato al link
- XPointer: XML Pointer Language
  - Linguaggio per esprimere riferimenti a parti di un documento referenziato
    - #intro
    - #xpointer(id("intro"))
    - #xpointer(book/chapter[position() <= 5])

## XML schemas and XML Query

- XML schemas
  - Sostituiscono i DTD per specificare la struttura di un documento XML
  - La sintassi è identica a quella di un documento XML
  - E' possibile definire tipi di dati
- XML Query
  - Possibilità di accedere, interrogare e manipolare documenti XML
  - E' l'equivalente dell'SQL per i database relazionali

## HTTP - HyperText Transfer Protocol

- HTTP e' un protocollo ASCII che
  - definisce il formato dei messaggi scambiati tra client e server
  - le azioni da intraprendere alla ricezione
- Versione standard: HTTP 1.1 (RFC 2068)
- E' un protocollo stateless (senza stato)
  - il client contatta il server alla porta 80, usando TCP
  - richiesta del client al server
  - risposta del server al client
  - connessione TCP chiusa
- I cookies sono stati introdotti per realizzare applicazioni web state-based
- Le richieste e le risposte hanno intestazioni e corpi simili a SMTP e MIME
- HTTP daemon piu' diffuso: Apache (freeware) - <http://www.apache.org/>

## Esempio di richiesta/risposta HTTP

```
St7a62% telnet www.w3.org 80
```

```
Trying 18.23.0.22...
Connected to www.w3.org.
Escape character is '^]'.

```

```
HEAD /Protocols/ HTTP/1.0
```

```
HTTP/1.1 200 OK
Date: Thu, 23 Apr 1998 12:47:22 GMT
Server: Apache/1.2.6
Last-Modified: Tue, 21 Apr 1998 17:57:50 GMT
Content-Length: 21259
Accept-Ranges: bytes
Connection: close
Content-Type: text/html; charset=ISO-8859-1
```

## Richiesta HTTP

- Request header line
  - method azione da intraprendere (GET, HEAD, PUT, POST, ...)
  - identifier indirizzo della risorsa target (solo filename)
  - version versione http
- Request header fields
  - From: email address
  - User-Agent: quale browser (Mozilla/...)
  - Accept: lista di tipi/sottotipi MIME accettabili (text/html, image/gif, ...)
  - If-Modified-Since: Restituisci il documento se piu' recente
  - ...
- Entity Body
  - usato per passare al server informazioni non predefinite



# Risposta HTTP

- Response header line
  - version    http version
  - code        (200 OK, 400 bad request, 404 not found, ...)
- Response header fields
  - MIME Version:    versione MIME per codificare il messaggio
  - Server:            software usato per il server
  - Date:             data e tempo all'origine del messaggio
  - Content-type:    tipo/sottotipo MIME e codifica usato per il corpo
  - Content-length:  lunghezza in byte del corpo
  - ...
- Entity body
  - generalmente il documento HTML