

HTML 2° parte

(HyperText Markup Language – linguaggio descrittivo di ipertesti)

Augusto Scatolini (webmaster@comunecampagnano.it)

Ver. 1.0 Aprile 2011

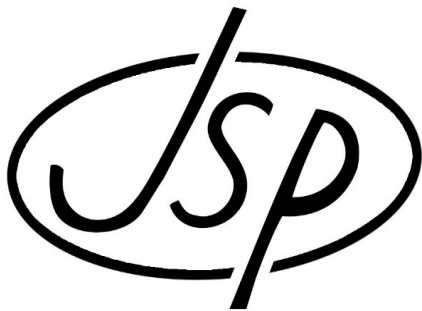
<!DOCTYPE
<HTML>
<HEAD>
<TITLE>RA
<LINK REV:
<META NAM

CSS

CGI



<?xml?>



ASP



PAGINE WEB STATICHE E DINAMICHE

PREMESSA

Dopo aver visto la guida introduttiva, generale sull'HTML, (linguaggio per creare pagine statiche) disponibile all'indirizzo <http://www.comunecampagnano.it/gnu/mini-howto/HTML/HTML.pdf> mi sembra opportuno completare il quadro generale del mondo web mostrando una carrellata di pagine e servizi che rendono il web dinamico.

*La maggior parte delle definizioni, riportate in questa guida, provengono direttamente da **Wikipedia**, (inutile inventare l'acqua calda). L'obiettivo di questo mini-howto è quello di creare un quadro complessivo di tutti i tipi di pagine web, dinamiche, che possono essere visualizzate su un browser.*

La locuzione **Web statico** viene utilizzata per identificare un paradigma di progettazione nel web che prevede un'interazione sostanzialmente unilaterale: l'utente può visualizzare i contenuti di un sito ma non modificare lo stato né le informazioni.

Il linguaggio di marcatura padre del web statico è l'**HTML**.

I siti web basati su questo paradigma sono stati la prima tipologia e per questa ragione estremamente diffusi. Basati su pagine web statiche, generalmente senza l'uso di **CSS**, permettono all'utente di navigare i contenuti, visualizzare le immagini e i dati, ma non elaborare alcunché né interagire dinamicamente (come si potrebbe invece fare con un *forum o una chat*).

Il **CSS** (**Cascading Style Sheets o Fogli di stile**) è un linguaggio, oltre a essere l'estensione del relativo file) usato per definire la formattazione di documenti HTML, **XHTML** e **XML**. Le regole per comporre il CSS sono contenute in un insieme di direttive (Recommendations) emanate a partire dal 1996 dal W3C. L'introduzione del CSS si è ritenuta opportuna per separare i contenuti dalla formattazione e permettere una programmazione più chiara e facile da utilizzare, sia per gli autori delle pagine HTML che per gli utenti.

I browser che interpretano il codice (X)HTML mostrano all'utente formattazioni predefinite per ogni tag che incontrano (così ad esempio i contenuti marcati con il tag <h1> avranno carattere 18pt e i contenuti marcati da <p> avranno carattere 12pt). Tuttavia questa formattazione è completamente sotto il controllo dell'utente, che può modificarla nelle Impostazioni del suo browser.

XML (sigla di **eXtensible Markup Language**) è un metalinguaggio di markup, ovvero un linguaggio marcatore che definisce un meccanismo sintattico che consente di estendere o controllare il significato di altri linguaggi marcatori. Costituisce il tentativo di produrre una **versione semplificata di SGML** che consenta di definire in modo semplice nuovi linguaggi di markup da usare in ambito web. Il nome indica quindi che si tratta di un linguaggio marcatore (markup language) estensibile (eXtensible) in quanto permette di creare tag personalizzati.

L'HTML tradizionale **non è un vero standard data la sua eccessiva flessibilità** ed il suo funzionamento immutato anche in **presenza di errori semantici, sintattici e grammaticali**. Ad affiancarlo c'è **XHTML**, ovvero l'**HTML tradizionale basato su XML**, con la sua struttura rigida e con le sue stesse regole.

Per permettere agli autori di definire l'aspetto delle loro pagine, dal 1993 in poi **Netscape Navigator** ed **Internet Explorer**, i due browser che si disputavano gli utenti nella nota guerra dei browser, presentarono tag proprietari, ovvero non aderenti agli standard e non compatibili con i browser concorrenti. Un esempio di questi tag è che va a definire il font.

Questi tag proprietari di formattazione erano l'unico modo per gli autori di definire la formattazione e così il loro uso è diventato massiccio.

Tuttavia questi tag presentano tre problemi:

1. Il primo problema è costituito dalla lunghezza di questi tag. Se confrontata con una pagina che adotta il linguaggio CSS, una pagina che non lo adotta è in genere più pesante (in termini di bit) in un rapporto che spesso raggiunge il 200%. Inoltre **le istruzioni CSS possono essere raccolte in un file esterno che rimane memorizzato nella cache del browser**, riducendo ulteriormente la quantità di dati che i server devono trasmettere.

2. Il secondo problema risiede nella mancanza di logica del codice (X)HTML. Un codice non aderente agli standard, ridondante e confuso comporta infatti molto lavoro aggiuntivo per i browser, che devono cercare di **correggere ed interpretare (quando possibile) direttive arbitrarie**.

3. Il terzo problema comincia a diventare sempre più rilevante ed è la **mancanza di compatibilità con i nuovi computer palmari e gli smartphone**. Queste pagine infatti sono progettate per schermi con risoluzione minima 800x600 pixel. I palmari, che hanno una risoluzione inferiore ed una forma dello schermo ben diversa dal rapporto 4:3 dei monitor per computer, si trovano quindi impossibilitati a visualizzare correttamente la pagina e l'utente dovrà tentare di "decodificarla", operazione spesso molto scomoda.

Si tende ad evidenziare anche un'ulteriore questione, nelle pagine web non standard, ovvero l'uso del tag <table> (le tabelle) per realizzare l'impaginazione delle pagine web. Questo viene considerato dai puristi come inaccettabile in quanto le tabelle sono pensate per impaginare dati tabulari e non layout web.

Oltre a questo, l'uso delle tabelle comporta anche svantaggi pratici: aumento del peso delle pagine, come già indicato al problema 1, e peggioramento dell'accessibilità, in quanto l'ordine logico dei dati può dover essere sacrificato per motivi di layout. Nonostante tutto, alcuni **sostengono l'uso delle tabelle per la loro maggior facilità di implementazione**.

Una pagina web è formata fondamentalmente da due elementi: i contenuti veri e propri e la formattazione ovvero l'aspetto con cui i contenuti sono mostrati all'utente.

Il linguaggio HTML (e la sua evoluzione XHTML) ha come scopo quello di gestire i contenuti associandone o specificandone allo stesso tempo la struttura grafica (layout) all'interno della pagina web da realizzare grazie all'utilizzo di tag diversi. Ogni tag (ad esempio <h1> o <p>) specifica un diverso ruolo dei contenuti che esso contrassegna.

Il termine **WEB dinamico** viene utilizzato per indicare tutte quelle applicazioni Web che interagiscono attivamente con l'utente e che **modificano le informazioni mostrate** all'utente in base alle informazioni ricevute dall'utente stesso. Esso si contrappone al Web statico che rappresenta il vecchio paradigma di programmazione Web.

Esempi tipici di applicazioni web dinamiche sono i siti di **e-commerce** che ovviamente devono interagire con l'utente ed adeguarsi alle sue scelte.

Molti siti web sono considerati **erroneamente "dinamici"** dato che consentono all'utente di personalizzare l'impaginazione o le informazioni mostrate o aggiornare in maniera efficiente (dinamica appunto) i contenuti modo da renderli più vicine alle richieste dell'utente o del programmatore web.

Altri siti web vengono **erroneamente considerati "dinamici"** semplicemente perché sono fatti con tecnologia Flash o perché contengono immagini animate.

Negli ultimi anni ha avuto un'elevata diffusione i siti di tipo **wiki**. In questi siti l'utente non solo interagisce con il sito per decidere quali informazioni visualizzare, ma a sua volta l'utente può fornire informazioni e contenuti se lo desidera. La stessa Wikipedia infatti è un sito web dinamico implementato attraverso un **CMS** scritto in **PHP**.

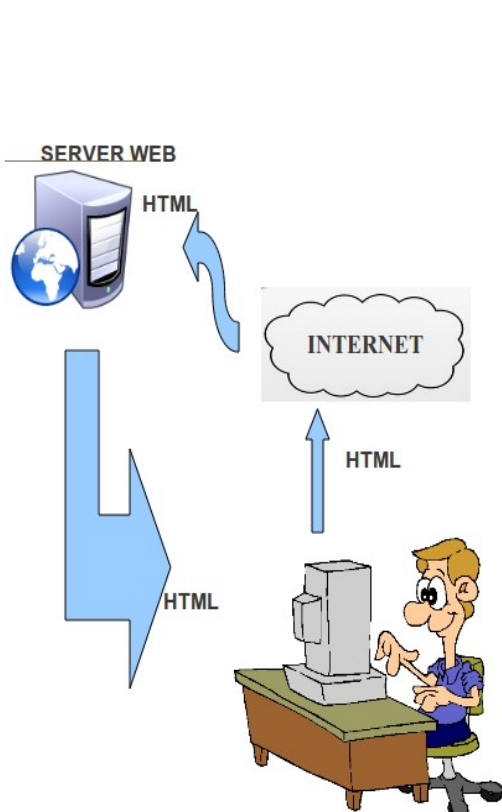
Una pagina web dinamica è una pagina web il cui contenuto è, in tutto o in parte, generato sul momento dal server e può essere quindi diversa ogni volta che viene richiamata.

Questo significa che non si utilizza direttamente il linguaggio HTML, ma che si ricorre a dei linguaggi di programmazione (i **linguaggi di scripting**) che si occupano della creazione della pagina nel momento in cui questa viene visitata, anche interagendo con i visitatori, e possono variare l'output HTML successivo dopo calcoli eseguiti con apposite strutture messe a disposizione dal linguaggio di scripting che si usa.

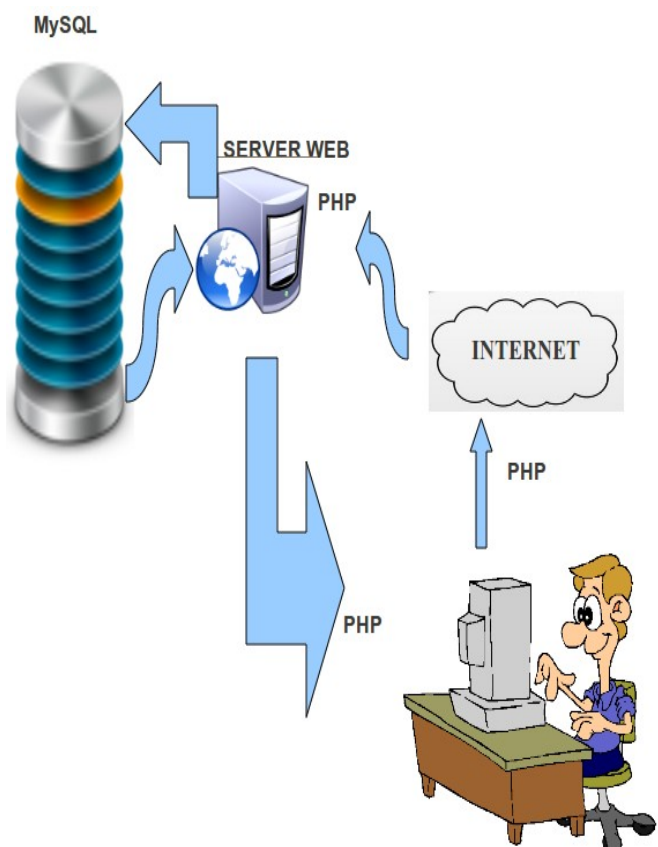
Alcuni esempi

Tutti i **motori di ricerca**, come ad esempio Google, sono costituiti interamente da pagine dinamiche. Inserendo una ricerca (ad esempio "pagine web dinamiche") verrà visualizzata una pagina generata al momento. I **Blog**, molto in voga negli ultimi anni, sono interamente basati su pagine dinamiche che accedono a basi di dati in cui vengono memorizzati gli interventi attraverso l'uso di piattaforme software dette CMS. I siti delle **banche** sono integralmente sviluppati in maniera dinamica in modo da poter accedere ai dati bancari degli utenti. Qualsiasi sito di **testata giornalistica** o di agenzia giornalistica si basa integralmente sulla generazione dinamica della pagina. Tutti i siti di **e-commerce**, basati su database per gli articoli e per le anagrafiche dei clienti, sono integralmente sviluppati in maniera dinamica.

WEB STATICO



WEB DINAMICO



Se digitiamo “<http://www.repubblica.it>” sulla barra degli indirizzi di un browser apparirà la seguente pagina web (la home page)



Se digitiamo “<http://www.digitpa.gov.it>” sulla barra degli indirizzi di un browser apparirà la seguente pagina web (la home page)



Ma quale pagina (quale nome.estensione) abbiamo chiesto al browser?

In realtà la prima domanda da farsi è “**ma come fa il browser ha trovare il sito e la relativa home page?**” quando digitiamo [repubblica.it](http://www.repubblica.it) o [digitpa.gov.it](http://www.digitpa.gov.it) ?

In realtà il browser interroga il Server dei Nomi di Dominio **DNS** che è stato settato dall'utente nei parametri di connessione, nel mio caso i DNS di **OpenDns** (o quelli di **Telecom** se non sono stati modificati). La domanda che il browser pone al server DNS è “**quale è l'indirizzo IP del dominio [repubblica.it](http://www.repubblica.it)?**”

Il server DNS cerca nel suo database e trova che il dominio [repubblica.it](http://www.repubblica.it) risponde all'indirizzo IP **231.92.16.101** mentre quello di [digitpa.gov.it](http://www.digitpa.gov.it) risponde all'indirizzo IP **67.215.65.132**

Questo è facilmente verificabile aprendo un terminale su sistemi GNU/Linux o una finestra DOS su sistemi Windows e lanciando un semplice comando: **nslookup** e a seguire il nome del dominio da identificare. Oppure tramite il comando ping.

```
File Modifica Visualizza Terminale Aiuto
augusto@augusto-desktop:~$ nslookup
> repubblica.it
Server:                208.67.222.222
Address:               208.67.222.222#53

Non-authoritative answer:
Name:   repubblica.it
Address: 213.92.16.101
> digitpa.gov.it
Server:                208.67.222.222
Address:               208.67.222.222#53

Non-authoritative answer:
Name:   digitpa.gov.it
Address: 67.215.65.132
>
```

Ora, sarebbe sufficiente digitare l'indirizzo IP (i 4 numeri) al posto del nome (sul browser) per aprire la home page del sito desiderato. Questo è vero per siti “relativamente semplici” che non sono composti da una complessa ragnatela di domini e **sottodomini per i quali non c'è corrispondenza univoca tra URL e indirizzo IP**. Infatti digitando l'indirizzo IP di [repubblica.it](http://www.repubblica.it) invece del sito di [repubblica](http://www.repubblica.it) si apre il sito de Kataweb (che è dello stesso gruppo editoriale). Se invece si prova ad aprire l'indirizzo IP di [digitpa.gov.it](http://www.digitpa.gov.it) non si apre proprio niente perché questo è l'indirizzo di un sottodominio (infatti nel nome ci sono due punti).

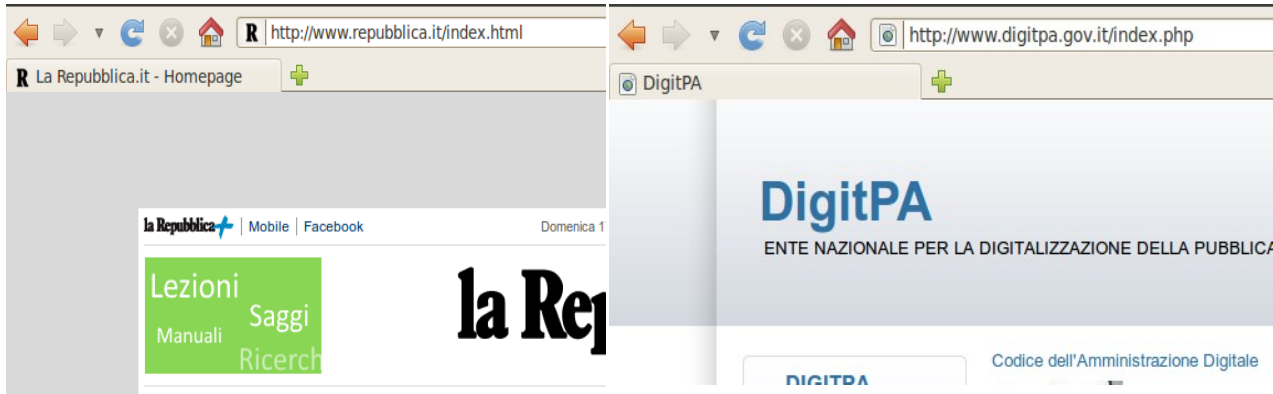
Tutti i siti governativi italiani sono del tipo nome.gov.it cioè di **terzo livello** (il primo livello è it, il secondo è gov). Il sito <http://www.gov.it> (di secondo livello), dato che è inutile, non è neppure visitabile con un browser.

Un sito come **aruba.it** che ha indirizzo IP **62.149.188.154** invece è interrogabile digitando sia <http://62.149.188.154> che <http://www.aruba.it>

Torniamo alle pagine

Una volta che il server DNS ha rintracciato l'indirizzo IP di `repubblica.it` cosa succede? Succede che il server WEB (dove risiede il sito `repubblica.it`) va a verificare nelle sue impostazioni (dove è settato il nome della home page) e mostra il contenuto della pagina `index.html`

Analogamente quando digitiamo `digitpa.gov.it` il web server verifica nelle sue impostazioni e mostrerà la sua home page, ma in questo caso sarà la pagina `index.php`



MA QUANTI TIPI DI PAGINE ESISTONO?

CGI

Common Gateway Interface (acronimo CGI, in italiano: interfaccia comune, nel senso di standard, per gateway), è una tecnologia standard usata dai web server per interfacciarsi con applicazioni esterne.

Ogni volta che un client richiede al web-server un URL corrispondente ad un documento in puro HTML gli viene restituito un documento statico (come un file di testo); se l'URL corrisponde invece ad un **programma CGI**, il server lo esegue in tempo reale, generando dinamicamente informazioni.

Il CGI è la prima forma di elaborazione lato server implementata. Quando ad un web server arriva la richiesta di un documento CGI (solitamente con estensione `.cgi`, `.exe` o `.pl`) il server esegue il programma richiesto e al termine invia al browser l'output del programma.

La tendenza, comunque, è di abbandonare le applicazioni CGI per diversi motivi. Uno dei principali è che i CGI sono solitamente in **forma compilata** e non di sorgente e questo comporta la difficile verifica della bontà dell'applicazione stessa rendendo il server che le esegue **vulnerabile ad attacchi** di tipo buffer overflow.

JSP

JSP significa **Java Server Pages** e serve per programmare lato-server. Le JSP non hanno nulla a che vedere con JavaScript. Sono capaci di produrre in maniera semplice e veloce del contenuto web dinamico. Le **servlet** sono oggetti che operano all'interno di un server per applicazioni (per esempio, Tomcat) e potenziano le sue funzionalità.

La parola servlet deriva da una precedente, **applet**, che si riferisce a piccoli programmi scritti in linguaggio Java che si eseguono all'interno di un browser lato client.

Per contrapposizione, una servlet è invece un programma che si esegue in un server web.

Sebbene molti siti continuino ad utilizzare il precedente linguaggio **JHTML**, questa tecnologia è stata ampiamente superata da JavaServer Pages.

PHP

PHP (acronimo ricorsivo di "**PHP: Hypertext Preprocessor**", preprocessore di ipertesti) è un linguaggio di scripting interpretato, con licenza open source e libera (ma incompatibile con la GPL), originariamente concepito per la programmazione Web ovvero la realizzazione di pagine web dinamiche.

Attualmente è utilizzato principalmente per sviluppare applicazioni web lato server ma può essere usato anche per scrivere script a riga di comando o applicazioni standalone con interfaccia grafica.

Viene comunemente utilizzato per interrogare (query) database quale **MySQL**

ASP

Le **Active Server Pages** (Pagine Server Active, in genere abbreviato in ASP) sono pagine web contenenti, oltre al puro codice HTML, degli script che verranno eseguiti dal server per generare runtime il codice HTML da inviare al browser dell'utente (proprio per questo vengono in genere definite pagine web dinamiche).

In questo modo è possibile mostrare contenuti dinamici (ad esempio **estratti da database** <**tipicamente MDB di MS Access**> che risiedono sul server web) e modificarne l'aspetto secondo le regole programmate negli script, il tutto senza dover inviare il codice del programma all'utente finale (**al quale va inviato solo il risultato**), con un notevole risparmio di tempi e banda.

I linguaggi utilizzati sono **VBScript** e **JScript** per l'ambiente ASP e **VB.NET, C# e J#** (il simbolo # si legge sharp) per l'ambiente **ASP.NET** (anche se il più utilizzato è proprio VB.NET).

Grazie a questi linguaggi il sistema dinamico può comunicare lato server con tutti gli oggetti presenti sul sistema, infatti le possibilità offerte dal sistema sono fortemente orientate verso l'interfaccia con un corrispondente database, rendendo così possibile lo sviluppo di siti dinamici basati sulle informazioni contenute nel database.

È possibile interfacciare le pagine ASP con qualsiasi tipo di database che abbia un driver OLE-db o ODBC, come ad esempio **Access, SQL Server, MySQL, Oracle, Firebird, Sybase** e tanti altri.

ASPX

ASP.NET è un insieme di tecnologie di sviluppo di software per il web, commercializzate dalla Microsoft. Utilizzando queste tecnologie gli sviluppatori possono realizzare applicazioni Web e servizi Web (Web Service).

La vasta disponibilità di controlli, classi e strumenti di sviluppo consente di ridurre sensibilmente i tempi di sviluppo dei moduli software, e mette lo sviluppatore in condizione di riutilizzare facilmente uno stesso codice in applicazioni differenti.

ASP.NET ha inoltre semplificato significativamente le funzionalità di accesso ai dati rispetto ad ASP. Ad esempio, è molto più semplice ed immediato generare una pagina contenente una lista di dati prelevati da un database.

CFM

Il linguaggio **CFML (ColdFusion Markup Language)** è un linguaggio di scripting fortemente orientato all'accesso ai database.

Molto apprezzato nello sviluppo di applicazioni enterprise, non ha goduto di grande supporto dagli sviluppatori poiché **l'interprete lato server non è gratuito come per i linguaggi più diffusi**.

ColdFusion è una tecnologia server creata da **Allaire**, ora distribuita da **Adobe**, che elabora pagine con l'estensione **.cfm e .cfml**.

Nella versione 9 (ottobre 2009) vengono migliorate la possibilità di inserire componenti **AJAX** e l'integrazione con le piattaforme **Adobe Flash**, **Adobe Flex** e **Adobe AIR**, grazie all'inclusione dello strumento Flash Builder 4 all'interno di ColdFusion Builder.

Riassumendo, sul web si possono incontrare le seguenti estensioni:

file.htm
file.html
file.css
file.xml
file.cgi
file.jsp
file.php
file.asp
file.aspx
file.cfm
file.cfml

ne avrò dimenticata qualcuna, sicuramente!

FINE

Questo documento è rilasciato con licenza Copyleft
(tutti i rovesci sono riservati)

altre miniguide su
<http://www.comunecampagnano.it/gnu/miniguide.htm>
oppure direttamente su
<http://miniguide.tk>