

Architettura dell'Informazione

2. Introduzione a Internet e al WWW

Paolo Milazzo

Dipartimento di Informatica, Università di Pisa

<http://www.di.unipi.it/~milazzo>

milazzo@di.unipi.it

Master in ICT e Turismo

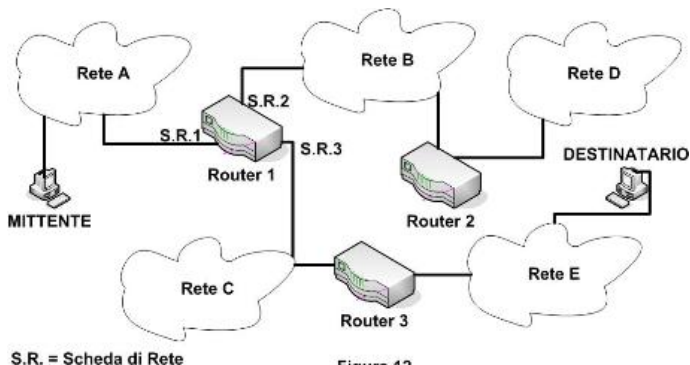
A.A. 2015/2016

Che cos'è Internet

- Internet è una **rete di reti**
- **Le singole reti hanno caratteristiche non omogenee:**
 - ▶ diversi protocolli di comunicazione (interni alle singole reti)
 - ▶ diversi tipi di elaboratori (PC, server, dispositivi mobili, router,)
 - ▶ diverse convenzioni nella codifica delle informazioni (e.g. il carattere di “return” nel testo)

Che cos'è una rete di reti

- Una **rete di reti** è il risultato del collegamento di un numero di reti (anche diverse tra loro) tramite dispositivi detti **router** (o **gateway**)
- Un modem ADSL è un esempio di router molto semplice e collega una piccola rete domestica (o aziendale) alla rete del fornitore del servizio ADSL (detto ISP - Internet Service Provider)
- I messaggi da un nodo di una rete (e.g. un PC) a un nodo di un'altra rete vengono **instradati** dai router e attraversano le reti intermedie



Che cos'è Internet

- Internet è una **rete di reti su scala globale**
- L'infrastruttura di Internet è basata su tecnologie eterogenee (linee telefoniche, fibre ottiche collegamenti satellitari, ecc...)
- La comunicazione tra due nodi A e B su Internet (o in generale su una rete di reti) avviene in 3 passi:
 - 1 Comunicazione tra A il router della propria rete (router 1)
 - 2 Comunicazione tra il router 1 e il router della rete di B (router 2) eventualmente mediata da altri router lungo il percorso
 - 3 Comunicazione tra il router 2 e B
- Tutte queste comunicazioni consistono in scambi di messaggi digitali la cui forma è stabilita da specifici **protocolli di comunicazione**

Protocolli di comunicazione (1)

- Il concetto di **protocollo di comunicazione** è essenziale nell'ambito del funzionamento reti
- Un protocollo di comunicazione stabilisce:
 - ▶ le **regole** da seguire pedissequamente nello scambio dei messaggi
 - ▶ il **formato** dei messaggi da scambiare

Protocolli di comunicazione (2)

Esempi di protocolli di comunicazione dalla vita quotidiana:

- Chiamata telefonica:

- ▶ **Regole da seguire:** comporre il numero, ascoltare il segnale di libero o occupato, parlare, riagganciare...
- ▶ **Formato dei messaggi:** messaggi vocali, iniziare dicendo “pronto”

- Invio di una lettera:

- ▶ **Regole da seguire:** scrivere la lettera, inserirla in una busta, scrivere l'indirizzo del destinatario, affrancare, inserire nella cassetta della posta...
- ▶ **Formato dei messaggi:** fogli di carta, intestazione “Caro Tizio”, conclusione “Distinti Saluti”,

- Scambio di bigliettini a scuola:

- ▶ **Regole da seguire:** attirare l'attenzione del destinatario, attendere che il professore non veda, passare furtivamente il messaggio, ...
- ▶ **Formato dei messaggi:** piccoli biglietti di carta, scrittura compatta, biglietto ben piegato,....

Protocolli di comunicazione (3)

- I protocolli di comunicazione possono essere a diversi **livelli**.
- Ad esempio:
 - ▶ Il protocollo di chiamata telefonica assume che tra i telefoni dei comunicanti esista un protocollo di comunicazione “di più basso livello” che stabilisce: come i segnali elettrici vengono inviati da un telefono all'altro, come devono essere tradotti nel segnale libero/occupato, cosa deve accadere quando un utente riaggancia il telefono,
 - ▶ Il protocollo di invio di una lettera assume che “a livello più basso” esista un protocollo di distribuzione delle lettere che fa sì che ogni lettera venga raccolta, raggiunga l'ufficio postale più vicino, poi venga inviata all'ufficio postale del destinatario e infine venga recapitata.

Protocolli di comunicazione (4)

- Analogamente i protocolli di comunicazione delle reti di computer sono realizzati a livelli:
 - ▶ I protocolli di basso livello gestiscono le connessioni fisiche ed elettromagnetiche tra i nodi delle varie reti
 - ▶ Protocolli di livello più alto gestiscono lo scambio di sequenze di segnali digitali (pacchetti) tra i nodi
 - ▶ Protocolli di livelli ancora più alto gestiscono lo scambio di messaggi tra le applicazioni che vengono eseguite sui vari nodi (e.g. messaggi di posta elettronica)

Protocolli di comunicazione (5)

- La strategia che di solito si utilizza per collegare protocolli di comunicazione su livelli differenti è **l'incapsulamento** dei messaggi
 - ▶ un messaggio di un protocollo di livello alto viene inserito nel corpo del messaggio del livello più basso
 - ▶ è come mettere una busta dentro l'altra (meccanismo delle scatole cinesi)



Protocolli di comunicazione (6)

- Primo esempio di protocolli di comunicazione basati su livelli e incapsulamento:
 - 1 Impero-A e Impero-B sono due imperi in guerra
 - 2 L'imperatore A vuole proporre all'imperatore B di fare la pace, scrive la richiesta su un foglio e lo consegna al generale dell'esercito A
 - 3 Il generale mette il foglio in una busta su cui scrive "per l'imperatore B – SEGRETISSIMO", e consegna la busta a un messaggero
(incapsulamento)
 - 4 Al messaggero viene assegnata una scorta armata (incapsulamento)
 - 5 La scorta conduce il messaggero sano e salvo nel territorio dell'Impero-B (trasmissione)
 - 6 Arrivati a destinazione, il messaggero consegna la lettera a un generale dell'esercito B e se ne va con la scorta (decapsulamento)
 - 7 Il generale apre la busta, la butta via e consegna il contenuto all'imperatore B (decapsulamento)
 - 8 L'imperatore B legge il messaggio

Protocolli di comunicazione (7) I

- Secondo esempio di protocolli di comunicazione basati su livelli e incapsulamento:
 - 1 Alice vuole mandare una email a Bob (bob@acme.it). Scrive il messaggio usando un programma apposito e clicca su “Invia”.
 - 2 Tramite l'indirizzo email il programma ottiene un codice numerico (indirizzo IP) che identifica il server che gestisce la posta elettronica della ACME. Costruisce quindi un “pacchetto” digitale che contiene il messaggio ed è etichettato con l'indirizzo IP (**incapsulamento**)
 - 3 Un driver del sistema operativo del computer di Alice si procura il numero di porta TCP del programma che gestisce la posta elettronica nel server del destinatario (per rendere l'idea... il numero di telefono dell'interno desiderato), prende il “pacchetto” e lo inserisce in una richiesta di connessione etichettata con il numero di porta TCP (**incapsulamento**)
 - 4 La scheda di rete del computer di Alice traduce il “pacchetto” in una sequenza di segnali elettrici da mandare nella rete, con opportuni dati di instradamento che consentono a questi segnali di raggiungere il server voluto (**trasmissione**)

Protocolli di comunicazione (7) II

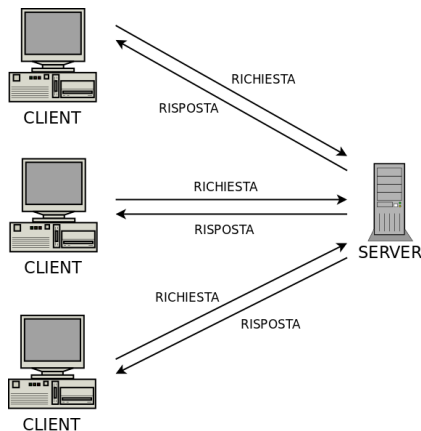
- 5 Il server ritraduce i segnali elettrici in informazioni digitali, scartando i dati di instradamento (decapsulamento)
- 6 Tramite il numero di porta TCP, un driver del sistema operativo del server identifica il programma a cui il pacchetto deve essere consegnato (che sarà il gestore delle mailbox e non, ad esempio, il programma che legge gli mp3...) (decapsulamento)
- 7 Il programma gestore delle mailbox mette l'email nella casella "Inbox" di Bob

I protocolli TCP/IP

- TCP e IP (che vedremo in seguito) sono i protocolli di comunicazione fondamentali su cui si basa Internet
- Le regole di scambio dei messaggi di questi protocolli fanno sì che la rete sia **robusta**
 - ▶ in grado cioè di funzionare in modo efficiente e affidabile anche se una parte di essa dovesse essere danneggiata

Architettura client-server (1)

- Al di sopra dei protocolli TCP/IP è possibile costruire applicazioni che realizzino servizi per gli utenti
- L'architettura principale che supporta i servizi messi a disposizione su Internet è l'**architettura client-server**



Architettura client-server (2)

- Un **server** è un programma che attende richieste che hanno la forma di messaggi che seguono un certo protocollo. Quando arriva una richiesta da un **client**, il server la analizza ed elabora una risposta da restituire al client. Un server, generalmente, può servire più client contemporaneamente
- Un **client** è un programma che si connette ad un **server**, fa una richiesta e attende una risposta
- Un client e un server possono essere eseguiti anche sullo stesso computer. Ad esempio: il sistema operativo (Windows, MacOS, Linux) si comporta spesso da server nei confronti dei programmi applicativi (Word, Excel,)

Architettura client-server (3)

- Nell'ambito di Internet le terminologie di client e server vengono spesso associate ai computer invece che ai programmi. Ad esempio, per computer server si intende il computer su cui è eseguito il programma server.
- Nell'ambito di Internet, richieste e risposte sono messaggi TCP/IP che viaggiano sulla rete.
- Esempio di client-server: il browser e il server web

Architettura peer-to-peer (P2P)

- Un'architettura alternativa a quella client-server che ha una certa diffusione su Internet è l'architettura **peer-to-peer (P2P)**
- L'architettura peer-to-peer assume che tutti i nodi coinvolti (computer o programmi) siano uguali, ossia che non ci sia un nodo che ha funzioni speciali (tipo il server)
- La gestione di una rete peer-to-peer è piuttosto complessa in quanto non c'è un'entità centrale di riferimento (il server)
- Questa architettura ha avuto particolare successo per applicazioni quali il file-sharing (condivisione di file) in cui non si vuole avere un'entità centrale di riferimento che si sobbarcherebbe la responsabilità legale del materiale condiviso (e.g. diritti d'autore e pirateria)

Protocolli applicativi

- I **protocolli applicativi** (utilizzati dalle applicazioni che realizzano servizi per gli utenti)
 - ▶ sono di livello più alto rispetto a TCP/IP (e si basano su essi)
 - ▶ definiscono le regole di comunicazione tra client e server per uno specifico servizio
- Esempi di protocolli applicativi
 - ▶ FTP (File Transfer Protocol): protocollo per il trasferimento di file
 - ▶ SMTP (Simple Mail Transfer Protocol): protocollo per il trasferimento dei messaggi di posta elettronica
 - ▶ HTTP (HyperText Transfer Protocol): protocollo per il trasferimento delle pagine web
 - ▶

Verso il World Wide Web

- A partire dalla metà degli anni '80, i numerosi diversi protocolli che sono stati inventati e sono entrati in uso su Internet avevano tutti differenti interfacce utente (programmi diversi per i diversi servizi)
- Per evitare una proliferazione di protocolli differenti e interfacce differenti per la diffusione dei dati e la comunicazione su Internet è stato proposto (e si è affermato) l'approccio basato su ipertesti
 - ▶ Un **ipertesto** è un documento elettronico che può includere contenuti multimediali (immagini, suoni,) e in cui si può passare da una pagina all'altra in maniera non sequenziale seguendo dei collegamenti (link)
 - ▶ Il **World Wide Web** è in realtà un unico grande ipertesto....

Sommario

1 Concetti introduttivi su Internet e sul World Wide Web

- Che cos'è Internet

2 Breve storia del World Wide Web

- La prima guerra dei browser
- La seconda guerra dei browser

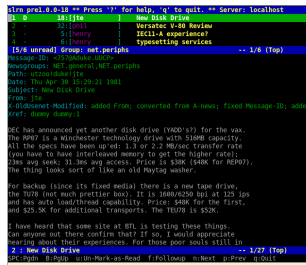
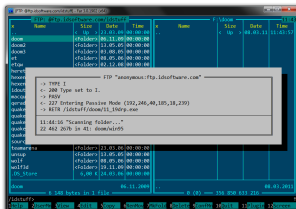
3 Che cos'è il World Wide Web?

Breve storia del World Wide Web (1)

- Il World Wide Web (WWW) nasce tra il 1989 e il 1991 come progetto del CERN di Ginevra affidato a un gruppo di ricercatori informatici tra i quali **Tim Berners-Lee** e **Robert Cailliau**.
- Lo scopo del progetto era la creazione di un sistema pratico ed efficiente per lo scambio di informazioni all'interno del CERN e tra i ricercatori del CERN e i loro collaboratori all'estero.
- I ricercatori identificarono due elementi chiave per la realizzazione di tale sistema:
 - ▶ **Internet**: la rete globale già esistente fin dagli anni '70
 - ▶ **Gli ipertesti**: documenti elettronici non organizzati in maniera sequenziale (come un libro) ma contenenti collegamenti (i **link**) che consentono di saltare da un punto all'altro nel documento

Internet prima del WWW:

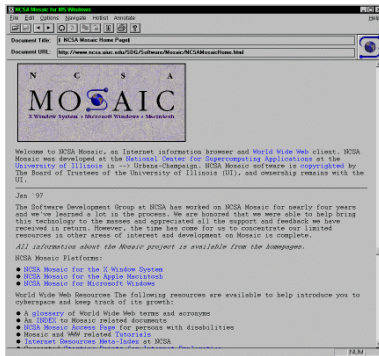
- strumenti diversi per ogni funzionalità (trasferimento file, consultazione notizie, invio messaggi,...)



Breve storia del World Wide Web (3)

Internet dopo il WWW:

- un unico strumento (l'ipertesto) per molte funzionalità (trasferimento file, consultazione notizie, invio messaggi,...)



Breve storia del World Wide Web (4)

- Nel 1991 i ricercatori realizzarono un prototipo basato su NeXT (una piattaforma hardware su cui girava una variante di Unix) già dotato di una sorta di interfaccia grafica per “navigare”.
- Alla fine del 1992 erano già disponibili 26 web servers in varie nazioni.
- Alla fine del 1993 i web server disponibili erano già oltre 200.
- Tra il 1992 e il 1993 il National Centre for Supercomputing Applications (NCSA) realizzò il primo web browser: chiamato **NCSA Mosaic**
- Nel 1994 Tim Berners Lee fonda il World Wide Web Consortium (**W3C**), un consorzio di istituzioni pubbliche e private (incluse le principali aziende informatiche) che si occupa di definizione e standardizzazione dei linguaggi e delle tecnologie utilizzate nell'ambito del Web.

Breve storia del World Wide Web (5)

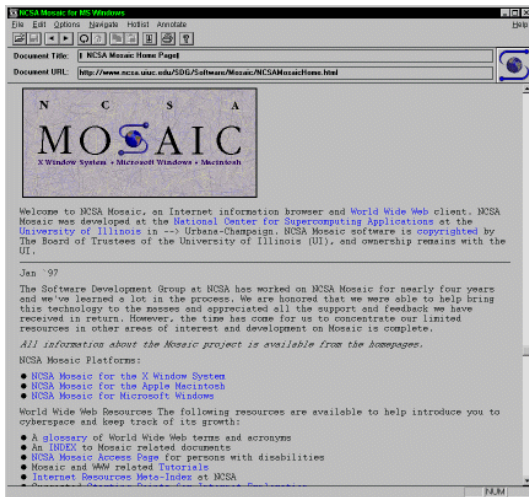
- In pochi anni il World Wide Web ebbe una crescita molto rapida, divenendo utilizzato da milioni di persone. Gli interessi commerciali attorno a tale successo portarono alla “prima guerra dei browser” (a cui ne seguirà una seconda attualmente in atto...).

La prima guerra dei browser (1)

- Nel 1994 Marc Anderseen (realizzatore di NCSA Mosaic) e Jim Clark (fondatore della Silicon Graphics) fondarono Netscape Corp. e realizzarono il browser Netscape Communicator basato su Mosaic
- Nel 1995 Microsoft lancia il browser Internet Explorer.
- Negli anni successivi (fino al 1998) i due browser combattono a colpi di nuove funzionalità, ma alla fine Netscape si dichiara sconfitta non riuscendo a competere con Internet Explorer (che è preinstallato su Windows...)
- Netscape Corp. viene venduta ad America On-Line (AOL) e il codice sorgente di Communicator viene reso pubblico creando una comunità Open Source (chiamata Mozilla.org) che ne porti avanti lo sviluppo.
- Nel 2000 esce una nuova versione del browser Mozilla, ricco di funzionalità ma molto lento.
- Nel 2004 la comunità di Mozilla riesce nell'intento di rivoluzionare il cuore del proprio browser e realizza Mozilla Firefox, un browser veloce, flessibile e multiplatforma che riscuote un rapido e crescente successo.

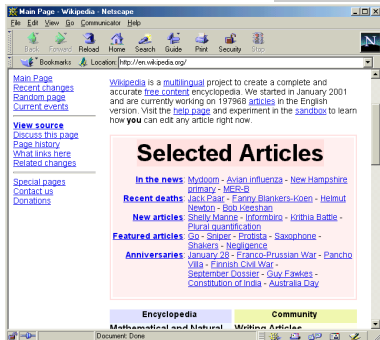
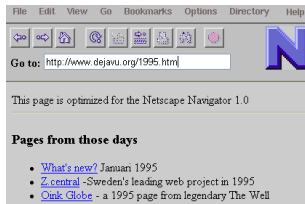
La prima guerra dei browser (2)

Il browser NCSA Mosaic



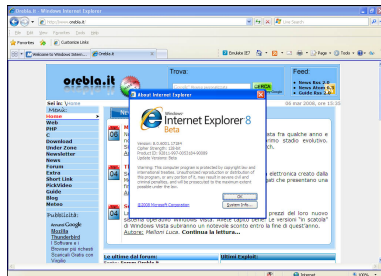
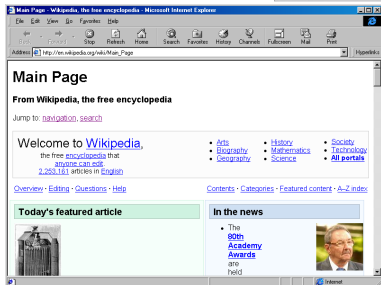
La guerra dei browser (3)

Alcuni esemplari di Netscape Communicator/Navigator (1.0, 4.7 e 9.0)



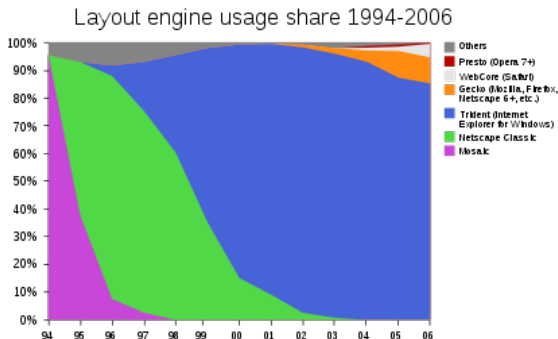
La prima guerra dei browser (4)

Alcuni esemplari di Internet Explorer (1.0, 4.0, 8.0)



La prima guerra dei browser (5)

Percentuali di utilizzo dei vari browser negli anni della prima guerra



Sommario

- 1 Concetti introduttivi su Internet e sul World Wide Web
 - Che cos'è Internet
- 2 Breve storia del World Wide Web
 - La prima guerra dei browser
 - La seconda guerra dei browser
- 3 Che cos'è il World Wide Web?

La seconda guerra dei browser (1)

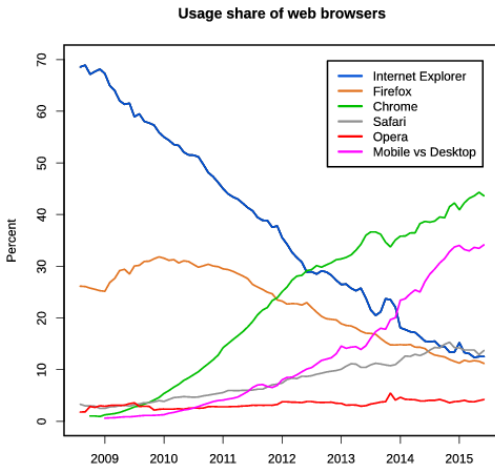
- Dopo aver vinto la prima guerra, Internet Explorer essendo in situazione di monopolio ha continuato il proprio sviluppo senza rispettare le direttive e gli standard proposti dal W3C (per prevenire la possibilità di concorrenza)
- Nel 2004/2006 due fattori importanti hanno riaperto la partita sui browser:
 - ▶ le numerose falle di sicurezza manifestate da Internet Explorer (che, essendo il browser più diffuso, era costantemente nel mirino degli hacker)
 - ▶ l'ingresso sul mercato di nuovi competitori di grosso calibro (oltre a Mozilla Firefox abbiamo Apple con il browser Safari e Google con il browser Chrome)
- Tutti i competitori di Internet Explorer supportavano adeguatamente gli standard del W3C.

La seconda guerra dei browser (2)

- Nel 2010, circa il 50% degli utenti usava Internet Explorer. L'altra metà usava browser aderenti agli standard (principalmente Mozilla Firefox)
 - ▶ In questo periodo sviluppare un sito web era piuttosto complicato. Bisognava effettuare molti test per garantire il funzionamento corretto sia su Internet Explorer che sugli altri browser...
- Con il passare del tempo la non-aderenza agli standard del W3C è diventato un fattore di svantaggio commerciale
 - ▶ Nelle ultime versioni Internet Explorer ha iniziato a supportare tutti gli standard allineandosi alla concorrenza
- Google è l'azienda che ha investito più risorse nella ricerca sulle nuove tecnologie per il web, dominando anche il lavoro del W3C
 - ▶ Questo fattore, unito alla grande forza commerciale e di distribuzione dell'azienda, hanno portato il browser Chrome a diventare il prodotto più utilizzato (e forse anche il più avanzato tecnologicamente)

La seconda guerra dei browser (3)

- Inoltre, negli ultimi anni sono nati anche altri browser spesso rivolti a specifici mercati (ad esempio per i dispositivi mobili) portando alla seguente distribuzione di utilizzo:



Sommario

1 Concetti introduttivi su Internet e sul World Wide Web

- Che cos'è Internet

2 Breve storia del World Wide Web

- La prima guerra dei browser
- La seconda guerra dei browser

3 Che cos'è il World Wide Web?

Che cos'è il World Wide Web?

Il World Wide Web è un “globally distributed hypermedia system”, ossia un ipertesto multimediale distribuito a livello mondiale:

- Un **ipertesto**, ossia un documento (elettronico) costituito da pagine consultabili in maniera non sequenziale attraverso l'uso di collegamenti (links) all'interno delle pagine stesse.
- **multimediale**, ossia che contiene testo, ma anche immagini, suoni, video, ecc...
- **distribuito a livello mondiale**, ossia tale per cui le pagine non sono tutte memorizzate in un unico computer, ma su innumerevoli server diversi. I collegamenti contenuti in una pagina di un server possono riferire pagine memorizzate su altri server.

Che cosa non è il World Wide Web?

Il World Wide Web **NON E'** Internet.

- Internet è la rete globale di computer su cui il World Wide Web viene diffuso.
- Internet si basa su protocolli di comunicazione che consentono di interconnettere reti diverse e basate su caratteristiche di connessione anche molto diverse tra loro
- Internet è il canale che consente di diffondere, oltre al World Wide Web, servizi quali la posta elettronica, il trasferimento di file (FTP), la messaggeria istantanea, ecc...

Le fondamenta del World Wide Web

I protocolli/linguaggi su cui si basa il World Wide Web (fin dalle sue origini) sono i seguenti:

- **Uniform Resource Locator (URL)** – Come riferire un documento
E' una convenzione su come scrivere i riferimenti ad altri documenti.
Viene utilizzato nei link contenuti nelle pagine web.
- **HyperText Transfer Protocol (HTTP)** – Come ottenere un documento
E' un protocollo di comunicazione che viene utilizzato per l'accesso a documenti ipertestuali via rete.
- **HyperText Markup Language (HTML)** – Il formato dei documenti
E' un linguaggio per la realizzazione di documenti ipertestuali

Estensioni del World Wide Web

A questi componenti di base si sono poi aggiunti nel tempo altri componenti che consentono di realizzare pagine web più evolute e dinamiche, ad esempio:

- **Cascading Style Sheet (CSS)**: un linguaggio per descrivere gli aspetti di presentazione di una pagina web (font, colori, ecc...)
- **Client-side Programming**: un insieme di linguaggi e tecnologie (e.g. JavaScript) per realizzare pagine più complesse e interattive che contengono del codice (i.e. programmi) che viene eseguito dal browser
- **Siti web dinamici (server-side programming)**: il server web non è più un semplice “memorizzatore di pagine” ma esegue applicazioni i cui output sono pagine web. Per questo nascono linguaggi quali PHP, ASP.NET,

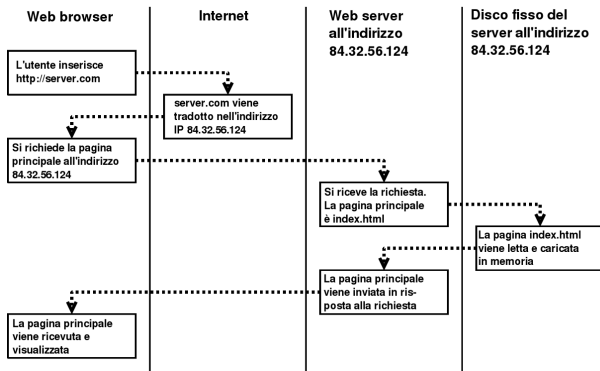
Architetture del World Wide Web

Alcune architetture comuni per la realizzazione di siti web sono le seguenti:

- **Sito web statico:** Ogni pagina del sito è memorizzata in un file diverso sul web server in formato HTML. Quando viene richiesta una pagina, il file corrispondente viene inviato al client e visualizzato così com'è dal browser.
- **Sito web con script:** Differisce rispetto all'approccio statico nel fatto che i file HTML possono contenere porzioni di codice (nel linguaggio JavaScript) che verranno eseguite dal browser per renderle maggiormente interattive (controllo sui dati immessi nella pagina, apertura di finestre di dialogo, ecc...)
- **Sito web dinamico:** I server web eseguono programmi che generano le pagine web richieste da un client. Tali programmi spesso interrogano un data-base per ottenere le informazioni che servono per costruire le parti principali delle pagine.

Architettura di un sito web statico (o con script)

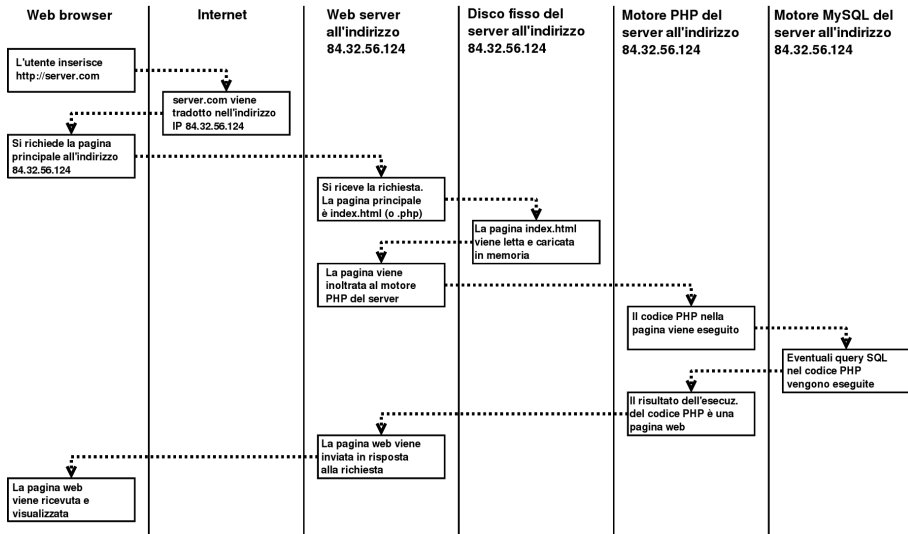
- Il web server ha un archivio di documenti HTML (eventualmente contenuti script) che invia al client su richiesta
- Il browser (eseguito sul client) visualizza il documento ricevuto (eseguendo eventuali script)



Architettura di un sito web dinamico (1)

- In un sito web dinamico il server web non ha un archivio di pagine, ma esegue un programma (web application) che **genera** le pagine su richiesta
 - ▶ Tra i linguaggi più usati per le web applications: PHP e ASP.NET, ma anche Java, Python, ...
- Comunemente la web application interagisce con un database usando opportune librerie ed il linguaggio SQL
 - ▶ Tra i database più usati: MySQL

Architettura di un sito web dinamico (2)



Nuove architetture di siti (e applicazioni) web

Negli ultimissimi anni la diffusione di collegamenti a banda larga e di dispositivi mobili (smartphone e tablet) ha portato alla nascita di nuove architetture di siti e applicazioni web:

- Applicazioni web che con interazione tra i programmi client-side e server-side (es. tecnologia AJAX originalmente applicata da Google per i suggerimenti automatici nella barra di ricerca)
- **Mobile web-applications**: siti web installabili sui dispositivi mobili come normali app
- Servizi **cloud**: risorse di memorizzazione (spazio disco) e di calcolo fornite tramite il web