



**Dipartimento
di Studi Umanistici**

Fondamenti di Informatica

Il World Wide Web

Giuseppe Cota

Laurea Triennale in Scienze e Tecnologie della Comunicazione - Università di Ferrara

In questa lezione

- 1 Il World Wide Web
 - La ricerca nel Web
- 2 Tecnologie Web
- 3 Web 2.0: Social Network, Smart Revolution e Internet of Things



Il World Wide Web



Il World Wide Web

- Uno dei più recenti e il più utilizzato servizio di Internet.
- Una *ragnatela* (web) di documenti multimediali collocati su computer (server) in tutto il mondo.
- **Killer application** di Internet
 - Killer application (applicazione assassina): un prodotto di successo che sfrutta una determinata tecnologia (quindi un'applicazione di quella tecnologia), grazie al quale quella tecnologia si impone nel mercato rispetto alle tecnologie concorrenti.
- Il più grande serbatoio di informazioni che sia mai esistito.
- Si accede tramite un **client** chiamato **web browser**
 - Microsoft Edge
 - Mozilla Firefox
 - Google Chrome
 - Apple Safari
- Tecnologie basilari: HTTP, HTML, URL.



Storia del Web

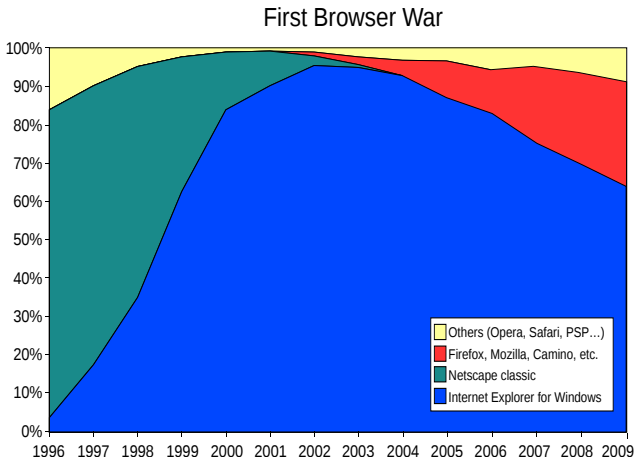
- **1980.** Un ricercatore del CERN, Tim Berners-Lee si rende conto che la lista delle sue attività, la rubrica telefonica e i suoi documenti si trovano su database diversi, su macchine diverse ed è complicatissimo accedervi simultaneamente.
- **1989.** T. Berners-Lee propone un primo progetto per la gestione documentale (Information Management: A Proposal). Tra le parole chiave contenute in questa proposta: simple protocol, hypertext, public domain software, network. Prime definizioni di HTTP, HTML e URL.
- **1990.** Il progetto viene rilanciato e per la prima volta viene utilizzato il termine World Wide Web per descrivere il sistema.
- **1991.** Viene rilasciato al pubblico il primo sito web basato sulle tecnologie HTTP, HTML e URL.
- **1993.** Le tecnologie HTTP, HTML e URL **diventano pubbliche.**
- **1993.** Viene rilasciato **Mosaic**, il primo web browser grafico.
- **1994.** Viene fondato il W3C (World Wide Web Consortium) da Tim Berners Lee (consorzio che definisce gli standard Web).



Storia del Web

Prima guerra dei browser

- **1995-2001.** Prima guerra dei browser.
 - Vincitore: Internet Explorer.

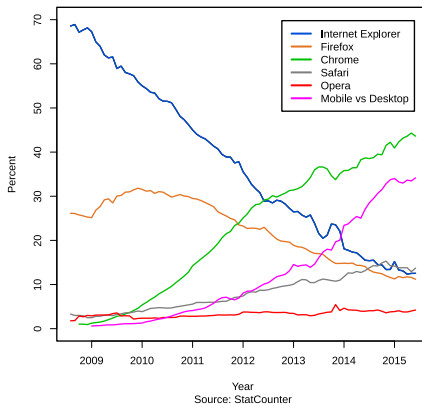


Storia del Web

Seconda guerra dei browser

- **2004-2017.** Seconda guerra dei browser.
 - Vincitore: Google Chrome.

Usage share of web browsers



Storia del Web

- **2004.** Tim O'Reilly fondatore della omonima e famosissima casa editrice, in una conferenza introduce il concetto di Web 2.0 riferendosi ad una evoluzione del Web introducendo servizi di “seconda generazione” come:
 - Social Network
 - Wiki
 - Blog



W3C: World Wide Web Consortium

- Il World Wide Web Consortium (W3C) viene fondato nel 1994 da Tim Berners-Lee (che tuttora ne è il direttore) come collaborazione fra il MIT (Boston) ed il CERN (Ginevra), con lo scopo di:
Condurre il World Wide Web al suo pieno potenziale sviluppando protocolli e linee guida che assicurino una crescita di lungo periodo del Web.
- Il W3C è un consorzio internazionale in cui le organizzazioni membro, lo staff permanente del consorzio ed il pubblico lavorano insieme per sviluppare gli standard del Web.



W3C: World Wide Web Consortium

Ogni standard, prima di diventare tale, cioè definitivo, subisce diversi passaggi evolutivi della sua definizione:

- **Working Draft (WD)**. Documento rilasciato dal W3C per essere rivisto dalla comunità.
- **Candidate Recommendation (CR)**. Documento che ha subito ampie revisioni e soddisfa i requisiti tecnici del gruppo di lavoro. Viene rilasciato per avviare la sperimentazione delle implementazioni.
- **Proposed Recommendation (PR)**. Documento maturo, con ampia revisione e verificate possibilità di implementazione in attesa dell'approvazione finale.
- **W3C Recommendation (REC)**. Documento di fatto uno standard per il W3C che ne raccomanda il massimo utilizzo e diffusione.
 - Esempio: HTML 5, CSS 2, ...

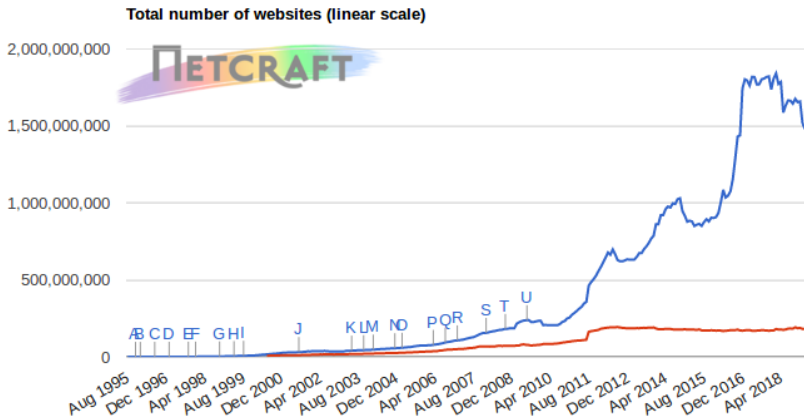


Il World Wide Web

- Il Web è costituito da **documenti** localizzati su calcolatori diversi collegati fra loro.
- Questi documenti sono consultabili tramite **web browser** (o semplicemente **browser**).
- **Sito web**: insieme dei documenti o pagine web gestiti da un'organizzazione o da un singolo autore, memorizzati su uno o più calcolatori.
- **Pagina web**: documento elettronico che può contenere:
 - testo
 - immagini, audio e video
 - link
- In definitiva, il Web è una rete costituita da miliardi di documenti interconnessi tra loro tramite **link ipertestuali** (o semplicemente **link**).
- Tecnologie basilari:
 - **HTTP, HyperText Transfer Protocol**: protocollo (livello applicazione) utilizzato dai computer per comunicare sul Web.
 - **HTML, HyperText Markup Language**: linguaggio che indica il formato che deve avere una pagina web.
 - **URL, Uniform Resource Locator**: usato per assegnare ad ogni documento un indirizzo univoco in modo da poterlo trovare sul Web.

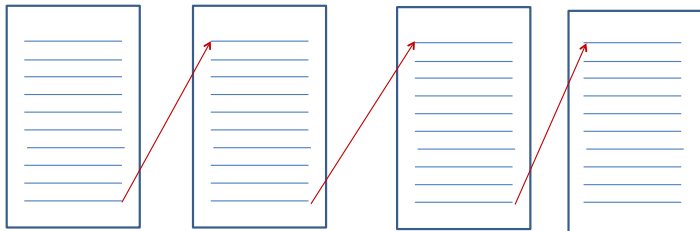
Quanti siti web esistono?

- **Febbraio 2019:** (Dati Netcraft)
 - 1 477 803 927 siti registrati
 - 181 081 966 siti attivi



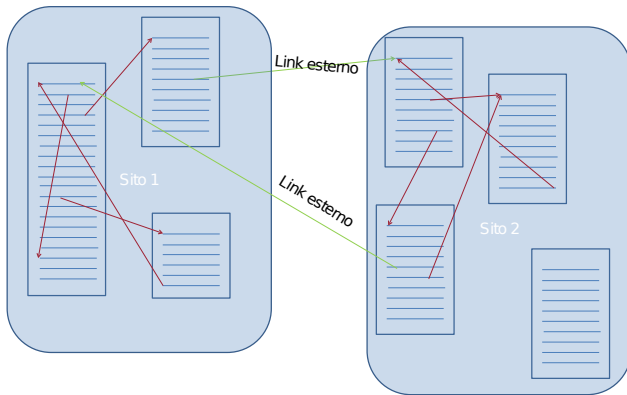
Letture sequenziale

- La lettura è sequenziale, il lettore legge dall'inizio alla fine pagina dopo pagina.



Iper testo e navigazione

- **Iper testo**: un documento (informatico) costituito da diversi **blocchi di testo**, collegati tra loro da nessi logici realizzati tramite **collegamenti (link)** che consentono al lettore di organizzare la lettura passando da un blocco all'altro (**navigazione**).



Tipi di link

- **Prima classificazione:**

- **Link interni:** collegamenti che puntano da una certa porzione di testo di una pagina web ad un'altra porzione di testo della stessa pagina.
- **Link esterni:** collegamenti che puntano ad un'altra pagina web.

- **Seconda classificazione:**

- **Link interni al sito:** collegamenti che puntano a pagine web dello stesso sito.
- **Link esterni al sito:** collegamenti che puntano a pagine web di altri siti.



Il browser

- Per utilizzare il web è necessario utilizzare un **(web) browser**.
- È un programma che ha le seguenti funzionalità:
 - Consente di inserire l'indirizzo del documento da visualizzare (l'URL) o di raggiungerlo attraverso i link.
 - Consente di sfogliare e navigare, avanti e indietro, le pagine già visitate.
 - Consente di stampare il documento visualizzato
 - Rende possibile il trasferimento (downloading) di file contenenti testo, video, audio, ecc. dal server al nostro computer.



Il World Wide Web

La ricerca nel Web



La ricerca nel Web

L'accesso ad una pagina web può avvenire in tre diverse modalità:

- L'utente può entrare **direttamente** digitando nella finestra di indirizzo del proprio browser l'indirizzo (l'URL) della pagina web.
- L'utente arriva **indirettamente** seguente un link presente in un altro sito.
- L'utente arriva in una pagina web attraverso un **motore di ricerca**.



I motori di ricerca

Le parti principali di un motore di ricerca sono due:

- **Crawler:** è quel programma che, in maniera invisibile all'utente, esamina i siti presenti nel web, spostandosi di collegamento in collegamento. Il crawler memorizza come parole chiave della pagina: i link che portano ad essa, il titolo della pagina e il testo contenuto.
- **Query Processor:** la ricerca viene svolta dall'utente attraverso il query processor (l'interfaccia).
 - 1 L'utente inserisce le parole chiave di interesse in una casella di testo.
 - 2 Dopo aver accettato le parole chiave, il query processor passa in rassegna tutte le pagine trovate dal crawler.
 - 3 Risponde con un elenco di link, ordinati secondo il suo criterio di pertinenza, a pagine che contengono le parole chiave inserite.



Quello che i motori di ricerca non sanno

- Se una pagina non è stata registrata da alcun crawler, allora fa parte di quello che viene chiamato **deep web**
- Esistono diversi motivi per cui una pagina non è stata registrata da un motore di ricerca.
 - Il Web cresce molto rapidamente e i crawler hanno troppo lavoro da fare. A volte sono necessarie settimane prima che una pagina web pubblicata sia rintracciabile da un motore di ricerca.
 - Alcune pagine non sono proprio trovabili:
 - Pagina senza testo, costituita solo da immagini, video o audio.
 - Pagina che non ha link verso altre pagine (link uscenti) né link che puntano ad essa (link entranti)
 - Il crawler non censisce volontariamente alcune pagine web perché hanno contenuto pericoloso o criminale.
- Il **dark web** è un sottoinsieme del deep web, le cui pagine web non sono raggiungibili tramite normale rete Internet, ma tramite reti sovrapposte ad Internet (**darknet**).
 - Esempi di darknet: Tor, I2P e Freenet.



Tecnologie Web



Tecnologie Web

Il Web si basa su tre standard principali:

- **HTTP, HyperText Transfer Protocol**: protocollo Internet del livello applicazioni sul quale si fonda lo scambio di informazioni nel Web. Prevede scambi di messaggi basati su comunicazione client/server.
- **HTML, HyperText Markup Language**: **linguaggio di marcatura (markup language)** con il quale si realizzano le pagine web. **Non è un linguaggio di programmazione**, ma un linguaggio che permette di arricchire i testi delle pagine web, di definire la formattazione, di definire i link e di indicare immagini, video e audio da visualizzare/riprodurre.
- **URL, Uniform Resource Locator**: stringa che identifica in maniera univoca una risorsa web (pagina web, immagine, ...).

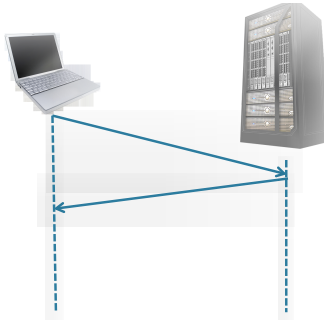


Modello client-server

- Il modello di comunicazione client-server è basato su due entità:

Client (Browser)

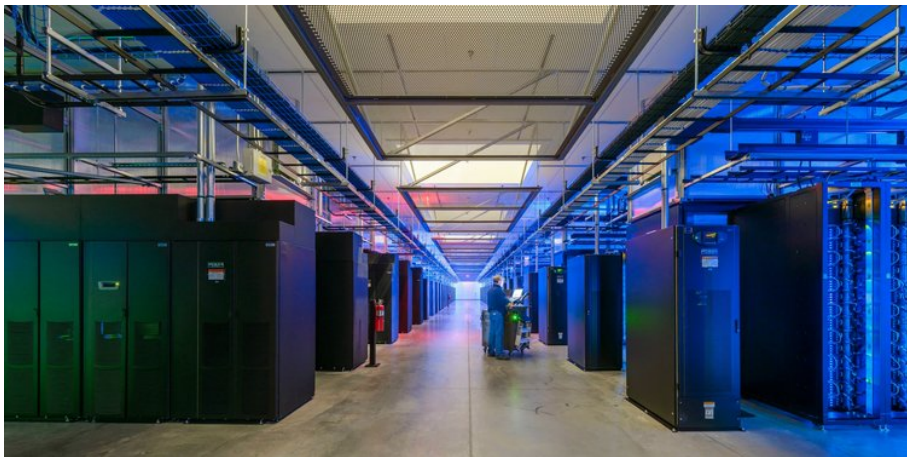
Effettua la richiesta di accesso (visita) di una pagina web al Web Server



Web Server

Risponde alla richiesta inviando un documento HTML

Server



URL, Uniform Resource Locator

- Formalismo che consente di identificare una risorsa web.
- Un URL è una stringa che ha il seguente schema:

protocollo://hostname:porta/path?query#frammento

- **protocollo**: protocollo che si sta usando. Esempio: http, https.
- **hostname**: indirizzo IP o nome di dominio associato almeno con un IP. Esempi: 193.204.255.20, www.unife.it, en.wikipedia.org (non è necessario che gli indirizzi web inizino con “www”).
- **porta**: (si vedano protocolli di livello trasporto). Per tradizione comunicazioni che usano il protocollo HTTP usano la porta 80, HTTPS usa solitamente la porta 443.
- **pathname**: percorso all'interno del filesystem del server. Schema: directory_1/.../directory_n/file.estensione
- **query**: informazioni parametriche.
- **frammento**: specifico punto in una pagina web.



Come si crea un sito Web

- Il linguaggio HTML è abbastanza semplice da imparare
- Creare una pagina web è facile, inoltre esistono numerosi strumenti che permettono di creare facilmente siti Web. Ad esempio: WordPress.
- Tuttavia, se si vuole creare un sito con pagine multimediali e numerose interazioni con l'utente, la faccenda si complica. Le grandi aziende si rivolgono ad altre aziende che si occupano dello sviluppo web.
- Una volta che il sito è stato creato, è necessario che venga pubblicato, cioè inserito nella memoria di un computer (server) collegato in rete che contenga un **server web**.
- Un **server web** è un programma che gestisce le richieste dei vari utenti e restituisce la pagine richieste.



Web 2.0: Social Network, Smart Revolution e Internet of Things



Web 2.0

- La definizione di Web 2.0 non è chiara e precisa, ma piuttosto sfuggente. Tuttavia è un termine che viene oramai utilizzato ovunque.
- Caratteristiche del **Web 1.0**
 - **Pagine web statiche.** Una volta visualizzata sul browser la pagina statica non cambia.
 - L'utente che naviga in rete è un consumatore passivo.
- Caratteristiche del **Web 2.0:**
 - **Pagine web dinamiche.** Una volta che è stata visualizzata sul browser la pagina, in seguito ad interazioni con l'utente, può cambiare.
 - L'utente non è solo consumatore ma pubblica contenuti ed informazioni (utente più partecipativo)
 - Social network, blog, wiki.



Social network

- Un **social network** è un raggruppamento di individui collegati tra loro da interessi e legami di vario tipo.
- I social network si suddividono in tre macro-settori:
 - professionale;
 - amicizia;
 - relazioni sentimentali.
- Hanno dato vita a nuove forme di comunicazione dette **virali**, che incoraggiano la rapida diffusione di un messaggio tramite condivisione tra utenti.
 - Non è necessario che il messaggio sia vero.



Smart revolution

- Con **Smart revolution** si intende la larga diffusione di dispositivi con accesso a Internet (e quindi alla modalità di accesso all'informazione), come tablet e smartphone, e i relativi mutamenti sociali, economici e politici.
- La rivoluzione smart ha già modificato le nostre abitudini, i nostri gesti, il modo di muoverci, la nostra cultura, le nostre scelte politiche,
- La rivoluzione smart ha cambiato i rapporti tra persone e **la comunicazione**.



Riferimenti principali

- Capitolo 5 di **Informatica Umanistica**. Marco Lazzari. 2a edizione. ISBN-13: 978-8838668555.
- Capitolo 9 di **Informatica di base**. Dennis P. Curtin D., Kim Foley, Kunal Sen e Cathleen Morin. ISBN-13: 978-8838615375.
- Slide Prof. Piva **Lezione 5 - Reti**. Corso di “Fondamenti di Informatica - Laurea Triennale in Scienze e Tecnologie della comunicazione” AA 2017-2018.



- Immagine slide 4. Wereon CC BY-SA 3.0
- Immagine slide 5. Usage share of web browsers (Source Net Applications).svg: arichnad, Daniel.Cardenas, Litehacker CC BY 3.0

