

The World Wide Web

- Il WEB è un *iperdocumento* costituito da *pagine* memorizzate su differenti nodi della rete internet
- WEB = ragnatela
- iperdocumento = documento con link (collegamenti)

Sviluppo del WEB

- Il WEB è il servizio INTERNET più diffuso, tanto che spesso i due termini vengono confusi
- Il WEB è divenuto un mezzo di comunicazione di importanza paragonabile a quella di Radio, TV e telefono

Un po' di storia

- Il progenitore del WEB moderno è stato un text-based Web browser scritto nel 1990 da Berners-Lee (un ricercatore del European Particle Physics Laboratory - CERN)
- Mosaic è il primo graphical Web browser e venne scritto nel 1993 dal NCSA (National Center for Supercomputing Applications)

URI

- URI = Universal Resource Identifier
Sintassi usata per identificare oggetti/risorse su Internet nel senso più generale possibile
- Gli URI consistono di 3 parti:
 - il nome del protocollo da utilizzare
 - il nome della macchina *host* della risorsa
 - il nome della risorsa (con il suo *path*)

URI: esempi

`ftp://ftp.cs.unibo.it/pub/techreports/`

- **ftp:**

La risorsa è accessibile tramite FTP

- **ftp.cs.unibo.it**

La risorsa è memorizzata sul server del dipartimento di Informatica

- **pub/techreports**

La risorsa è nella directory pub/techreports

URL

- URL = Universal Resource Locator
- Gli URL sono particolari URI che si riferiscono ad oggetti accessibili tramite protocolli standard esistenti
 - http http: ipertesti
 - ftp ftp: ipertesti
 - smtp mailto: posta elettronica

Browser

- I Browser sono programmi applicativi che:
 - dato un URL
 - scaricano via INTERNET la risorsa indicata
 - la decodificano e la *rendono* (visualizzano)
- Esempi:
 - Internet Explorer, Netscape Navigator

Come Navigare

- Il modo più semplice di navigare è:
 - indicare un URL di partenza
 - seguire i link
- Per cercare nuove pagine si usano **WEB SEARCH ENGINE** oppure **SUBJECT TREES**
- Sono archivi di URL che vengono aggiornati tramite “Spider” (che visitano la rete aggiornando gli archivi del motore di ricerca)

Search Engine

- Per attivare la ricerca si deve formulare una **search query** (interrogazione di ricerca): tipicamente vengono specificate le parole chiave della ricerca separate da spazi nel campo apposito, poi si avvia la ricerca. E’ possibile specificare un insieme di caratteri Jolly al posto di porzioni di parole chiave: ad esempio il carattere * (asterisco) serve a denotare qualsiasi cosa, per cui la ricerca del termine CALCO* potrebbe essere soddisfatta ad esempio dai termini CALCOLatore, CALCOLatrice, CALCOlo,

CALCOlare, CALCO, ecc. Allo stesso modo, il carattere + (o -) davanti a una parola chiave serve a specificare che tutti i riferimenti ottenuti dalla ricerca dovranno necessariamente includere la parola specificata (oppure dovranno necessariamente non includerla): es. +CALCO* -CALCOLO specifica che si cercano tutti e soli i documenti contenenti una qualunque parola che inizia con CALCO, a condizione che non appaia anche la parola CALCOLO.

- Il risultato è un elenco di *hit* (URL di pagine che soddisfano la richiesta)
- *Noise* (Rumore): fenomeno a causa del quale buona parte di hit non sono interessanti
 - Nel caso in cui vengano specificate più parole chiave per la ricerca, i riferimenti vengono ordinati ponendo per primi quelli che soddisfano meglio i criteri di ricerca, nel senso che le parole cercate appaiono in numero maggiore e possibilmente nei titoli. Nel caso in cui il numero di HIT sia molto elevato può essere opportuno filtrare maggiormente la ricerca, specificando parole chiave aggiuntive. In caso contrario, se non si sono trovati HIT, occorre ridurre il numero di parole chiave lasciando solo le più significative (ridurre il filtro).
 - È spesso possibile definire il target o la definizione della ricerca tra le seguenti categorie di oggetti: argomenti o categorie, siti web, pagine web, documenti
- Esempio: <http://www.altavista.com> e' un famoso motore di ricerca

Subject Tree

- Esiste un indice gerarchico di argomenti a partire dal quale ci si addentra nella ricerca sempre più raffinata del contesto o dell'argomento voluto.
- Seguendo questo indice si raggiunge un elenco di URL di pagine legate all'argomento considerato
- Esempio: <http://www.yahoo.com> è un motore di ricerca, ma mette anche immediatamente a disposizione il suo **subject tree**.
- NOTA: un elenco di search engine e subject tree è presente nelle pagine WEB del libro Internet 101