

Laboratorio di Informatica

Internet: servizi applicativi

WWW

World Wide Web: consultazione interattiva di ipermedia con modalità *point-and-click*

- browser e navigazione
- configurazione e sicurezza
- i motori di ricerca
- accesso ad altri servizi
 - posta elettronica
 - scaricamento di file
 - connessione remota
 - newsgroup

WWW: il protocollo

- Il protocollo di comunicazione usato dal web e che regola la trasmissione dei documenti ipertestuali si chiama HyperText Transfer Protocol (HTTP)
- Si basa sul protocollo TCP/IP
- I documenti devono però essere in uno specifico formato: HyperText Markup Language (HTML)

I browser

La guerra dei browser

- 1994: Netscape realizza importanti estensioni per HTML che solo il suo browser è in grado di gestire
 - differenti dimensioni e colori per i testi
 - fotografie, sfondi e immagini
 - pagine (frame) multiple in una singola finestra
- 1996: Netscape è il browser più popolare.
- Microsoft in risposta aggiunge estensioni non standard ad HTML leggibili solo da Explorer.

Verso la standardizzazione

- Il World Wide Web Consortium (W3C)

Connessione diretta ad Internet

- Per visitare un sito Web con un browser, si deve specificarne l'indirizzo (esplicitamente o selezionando un link)
- Il client invia la richiesta di connessione, formulata nel modo specificato da HTTP
- Ricevuta la richiesta, il server Web trasmette le informazioni al computer e il browser provvede a visualizzarle



Collegamento via modem

- Se siete collegati tramite un modem, allora le informazioni fra client e server passano attraverso il provider



Identificazione delle risorse

- Le risorse disponibili sulla rete (i documenti sui server) devono essere identificabili in modo univoco
- I browser identificano le risorse tramite indirizzi detti URL (**Uniform Resource Locator**)
- Un indirizzo URL ha la forma
`protocollo://server:porta/pathname`

Identificazione delle risorse

- **protocollo** indica il tipo di protocollo utilizzato
- **server** rappresenta l'indirizzo IP (numerico o simbolico) del computer a cui si vuole accedere
- **porta** identifica la porta a cui il protocollo fa riferimento
- **pathname** indica il percorso completo del file cercato

Identificazione delle risorse

- Alcune informazioni possono essere omesse (il server e/o il client sceglierà implicitamente dei valori predefiniti, detti di default)
 - il protocollo (per default `http`)
 - la porta (per default 80, quella associata a `http`)
 - il nome del file (per default `index.html`, `home.html` o altri) ma non il suo percorso

Identificazione delle risorse

- In altre parole i seguenti URL sono equivalenti
`http://laren.dsi.unimi.it:80/index.html`
`http://laren.dsi.unimi.it/index.html`
`http://laren.dsi.unimi.it:80`
`laren.dsi.unimi.it:80/index.html`
`http://laren.dsi.unimi.it`
`http://159.159.151.81`

Terminologia utile

- Utilizzando e configurando i browser ci si imbatte spesso in termini tecnici, di cui alcuni sono importanti anche per un uso corretto del web.
- proxy
- cookie
- applet
- javascript
- plug-in
- home page
- portale
- intranet
- extranet

I proxy

- Un proxy è un server (computer + programma) che svolge la funzione di "agente" per gli utenti di altri computer
- Configurando il proprio browser in modo che usi un proxy, le richieste di un documento non saranno fatte dal proprio computer direttamente al sito remoto, ma dal proxy server, che si preoccupa poi di fornirci il documento

Funzione dei proxy

- Il proxy dispone di una propria "cache", nella quale memorizza tutti i documenti recuperati negli ultimi tempi
 - se un utente richiede di caricare un documento che è già presente nella cache, il proxy provvede a spedirglielo direttamente senza contattare il sito remoto, e quindi in tempi più brevi
 - se il documento non è presente nella cache, il proxy contatta il sito remoto, recupera il documento, lo gira all'utente e lo memorizza nella cache

Cookie

- È un'informazione generata da un server Web e memorizzata nel calcolatore dell'utente, pronta per essere reperita in un accesso futuro.
- È un meccanismo per permettere la personalizzazione dalla parte dell'utente dell'informazione ottenuta dal Web.
 - Es. per personalizzare motori di ricerca con le proprie preferenze, per memorizzare liste della spesa, per permettere che un utente partecipi ad un concorso, ma solo una volta.

Cookie

- Quando un utente fa delle scelte, il server Web crea uno specifico *cookie*, che è essenzialmente una stringa di testo marcata contenente le preferenze dell'utente, e trasmette tale *cookie* al calcolatore dell'utente.
- Il browser Web dell'utente, se abilitato, riceve il *cookie* e lo memorizza in un file speciale chiamato lista di *cookie*.

Cookie e privacy

- Spesso non solo la memorizzazione di informazioni personali in un *cookie* avviene senza essere notata, ma anche l'accesso ad essa.
 - I server Web accedono automaticamente ai cookie ogni qual volta l'utente stabilisce una connessione per fare una richiesta.
- Questo meccanismo può facilmente essere usato per tenere conto delle scelte, numero di accessi, configurazione del calcolatore, ..., di un utente e pertanto lederne la privacy.

Cookie

I cookies sono memorizzati nella propria cartella personale in **Documents and Settings**. Un esempio di cookie è il seguente:

```
JEB245CB39D2F6E092E79F959761300487E7adtech.de/0363369  
13922988544892370355229522738*ADFC00010adtech.de/03  
6336913922988544892390355229522738*
```

Per verificare come questo meccanismo possa ledere la privacy di un utente, aprire una finestra di Internet Explorer e andare al sito

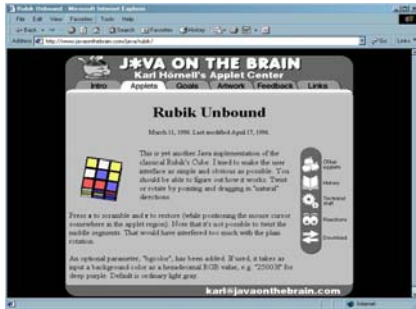
www.notrace.it

- Scegliere **Chi Sei?** nella scheda **Rubrica**
- Scegliere **Si** nella casella dello **Spy Test**
- Cliccare su **Entra**

Applet, JavaScript

- Consentono di inserire dei programmi all'interno di page web
 - **applet**: sono programmi eseguibili scritti in Java che possono essere eseguiti direttamente all'interno di un browser mediante l'interprete Java che si interfaccia con il browser.
 - **JavaScript**: è un linguaggio di scripting, il codice del programma è contenuto direttamente nella pagina web e viene interpretato ed eseguito dal browser al momento del caricamento del documento.

Applet



Esempi di applet Java sono reperibili al sito:

www.javaonthebrain.com/brain.html

Plug-in

- **Plug-in:** programma software (applicazione esterna) che estende le capacità di un browser in un modo specifico, ad esempio fornendo la possibilità di sentire un audio o vedere un video dal browser stesso.
 - Adobe Acrobat Reader
 - Apple QuickTime

Portale

- **Portale:** sito web che offre facile accesso attraverso link ad altri siti su internet.
- Ha la funzione di punto di partenza per navigare sul Web.
- Generalmente ha una collezione di siti e un motore di ricerca.
- I portali più recenti offrono anche la posta elettronica e altri servizi direttamente da una pagina che l'utente può personalizzare creando il proprio portale personale.

Portali generici

- Excite offre un portale personalizzato a **www.excite.com**
- Yahoo offre un portale personalizzato a **my.yahoo.com**
- Netscape offre un portale personalizzato a **my.netscape.com**
- Microsoft offre un portale personalizzato a **home.microsoft.com**
-

Portali specifici

Esistono portali specifici per:

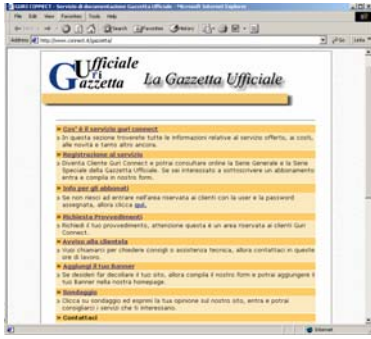
- Servizio di pagine gialle on-line **www.paginegialle.it**
- Trasporto ferroviario **www.trenitalia.com**
- Consultazione quotidiani **www.repubblica.it**
- Università **www.unimi.it**
- Ministeri **www.ministerosalute.it**

Intranet e extranet

- **Servizi Internet:** accessibili anche al di fuori della rete locale.
- **Servizi intranet:** accessibili solo internamente alla rete locale.
- **Servizi extranet:** accessibili anche al di fuori della rete locale, ma solo da parte di alcuni utenti riconosciuti e abilitati
 - es.: clienti e fornitori di un'azienda.

Esempio di Extranet

I servizi offerti dal sito della Gazzetta Ufficiale della Repubblica (www.connect.it/gazzetta) sono disponibili solo a utenti abbonati, previo pagamento di un canone.



I motori di ricerca su Web

- Permettono la ricerca di contenuti sul Web
- Problema di eccezionale difficoltà
 - milioni di pagine
 - breve ciclo di vita della pagina (circa 1,5 giorni)
 - inerente complessità di capire di che cosa parla una pagina / quant'è pertinente rispetto al concetto cercato

I motori di ricerca

- Un motore di ricerca è un server che fornisce un servizio di ricerca sulla rete per parole chiave
- Il motore di ricerca periodicamente guarda i documenti sulla rete e li indicizza in base ad delle parole chiave
- L'utente specifica delle parole chiave, e in risposta il motore di ricerca gli fornisce una lista di link ai documenti contenenti quella parola chiave

I motori di ricerca

- I motori di ricerca più famosi:
 - <http://www.google.com/>
 - <http://www.altavista.com/>
 - <http://www.yahoo.com/>
 - <http://www.excite.com/>
 - <http://www.lycos.it/>
 - <http://www.virgilio.it/>
 - <http://arianna.iol.it/>

I motori di ricerca

Strategie:




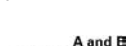
- Compilazione di digesti:
 - persone valutano la rilevanza della pagina rispetto a una tabella di contenuti-chiave che cattura i concetti più frequentemente cercati
 - a volte viene offerto un sunto del contenuto
 - esempio: www.yahoo.com
 - affidabile ma inevitabilmente poco dinamico

I motori di ricerca

- Indicizzazione:
 - le pagine vengono regolarmente copiate su dischi del server
 - la stringa passata dall'utente viene cercata esaustivamente (*string matching*)
 - esempio: www.altavista.com
 - più dinamico e completo
 - scarsa capacità di valutare la rilevanza della pagina rispetto al concetto cercato
 - # di occorrenze? Occorrenza nel titolo? # di occorrenze in pagine linkate? # di occorrenze in pagine che linkano?

I motori di ricerca

Operatori booleani

AND	Restringe il campo di ricerca racchiudere i termini presenti ma entrambi i termini presenti ma non necessariamente insieme <i>Freud AND Svevo</i>	
OR	Estende il campo di ricerca racchiudere i termini/espressioni uniti da OR tra () <i>Freud OR Svevo</i>	
AND NOT	Restringe il campo di ricerca Non è riconosciuto da tutti i motori <i>Freud AND NOT Svevo</i> Per escludere più di un termine, premettergli OR e racchiuderlo tra ()	
NEAR	simile a AND, però i due termini si devono trovare entro un certo numero di parole di distanza l'uno dall'altro <i>Freud NEAR Svevo</i>	

Laboratorio di Informatica Internet
Pagina 31

I motori di ricerca

Altri operatori

""	Racchiudere una frase o un'espressione tra "" "Freud e Svevo" il testo dovrà apparire identico
+	Restringe il campo di ricerca +Freud +Svevo entrambi in termini presenti ma non necessariamente insieme
-	Restringe il campo di ricerca +Freud -Svevo va premesso ai termini la cui occorrenza vogliamo escludere
*	Troncamento* (raro: AltaVista, NorthernLight e qualche directory) Svev* accetta tutte le terminazioni

Laboratorio di Informatica Internet
Pagina 32

Esercizio di Ricerca

Connettersi a un motore di ricerca

- <http://www.google.com/>
- <http://www.altavista.com/>
- <http://www.yahoo.com/>

E provare a ricercare ad esempio la frase
Dipartimento Chimica Milano
avvalendosi degli operatori

AND	OR	AND NOT	NEAR
+	-	""	*

Telnet

- Permette la connessione a calcolatori remoti.
- Il proprio calcolatore è impiegato come terminale a interfaccia alfanumerica.

```
Telnet - mercurio.trv.dsi.unimi.it
Dorothy Modifica Terminale 2

HP-UX mercurio 8.10.20 A 9000/021 (tty2)

login: norpurgo
Password:
Last successful login for norpurgo: Thu Nov 21 11:43:35 GMT 2002 on tty2
Last unsuccessful login for norpurgo: Tue Nov 19 14:20:04 GMT 2002 on tty2
Please wait...checking for disk quotas
```

Posta elettronica

- Sistema di messaggistica e scambio di file che riproduce alcuni aspetti del sistema postale
 - rende pressoché trasparente la comunicazione fra applicativi/reti/protocolli diversi
 - è forse il servizio di rete più popolare, anche per la facilità d'uso
- Protocollo SMTP
 - standard per il trasporto *host-to-host* della posta

Condizioni generali

Ogni utente ha un indirizzo

name@server.domain

- **name** può essere la login di un utente reale o di una lista di distribuzione
- **server**, generalmente, è un server su cui *gira* il programma di gestione della posta (es. mailstudenti)
- **domain** è l'indirizzo dell'organizzazione presso la quale si trova il server (es. unimi.it)
- Es.: **RM34765@mailstudenti.unimi.it**
- Per ogni parametro sono possibili *alias*

Formato del messaggio (SMTP)

- **Date:** <ora e data di spedizione>
- **From:** <login del mittente> (automatica)
- **To:** <uno o più indirizzi>
- **Subject:** <oggetto del msg>
- <una linea vuota>
- <il messaggio vero e proprio>
 - il messaggio è per definizione del protocollo SMTP solo in formato ASCII

Formato del messaggio (SMTP)

Altri campi dell'intestazione sono

- **Reply to:** indirizzo a cui inviare la risposta
- **Cc:** *carbon copy*, equivale a 'e. p.c.'
- **Encrypted:** rif. alla tecnica crittografica usata
- **Received:** gateway che l'ha inoltrato
- **Bcc:** non mostra gli indirizzi dei co-destinatari
- **Message-ID:** identificativo univoco del msg

Formato del messaggio (SMTP)

- **Attachments** (allegati):
 - immagini, testo formattato con word processor, file audio sono spediti come allegato (*attachment*) con una codifica ASCII *ad hoc* (MIME)
 - il file viene ricopiato in coda al messaggio ed arriva al destinatario con la stessa forma (nome, tipo, ecc.)
 - permette la distribuzione *mirata* ed *attiva* di novità, aggiornamenti ecc.
 - alternativa alla distribuzione *universale* (a meno di siti riservati) e *passiva* attraverso la pubblicazione sul web

Posta elettronica

- **Netiquette:** galateo per l'utilizzo di internet.
- **Spam:** E-mail non richiesta, generalmente di tipo commerciale, inviata a un grande numero di utenti.

Smilies

- :-) sto sorridendo
- :-(sono triste
- ;-) sto scherzando
- :-o sono sorpreso
- :-D sto ridendo

Software lato server

Il protocollo SMTP ha un'architettura *client-server*

- la parte *server* è il programma che si occupa di instradare e distribuire i messaggi
- è presente su alcune macchine che svolgono la funzione di *server* di posta elettronica
 - sendmail, comunemente in ambiente Unix/Linux
 - Outlook, comunemente in ambiente Microsoft

Software lato client

La parte *client* sul nostro computer è il programma che permette di

- **ricevere e leggere** sul nostro computer i messaggi destinati ai nostri indirizzi di posta
- **cancellare/salvare** i messaggi e gli eventuali attachments
- **rispondere** ai messaggi oppure ritrasmetterli per conoscenza
- **tenere alias** mnemonici degli indirizzi più usati e liste di distribuzione
 - Pine (Unix/Linux), Eudora (Win, Mac), il browser Navigator direttamente

L'ufficio postale

Il protocollo SMTP implementa alcuni servizi dell'ufficio postale

- sistema distributivo trasparente
- notifica di utente ignoto/mancata consegna
- effettiva lettura non garantita
- fermoposta: non perdiamo messaggi restando scollegati
- autenticazione del mittente non garantita (chi scrive?)

Protocolli di posta elettronica

- POP - Post Office Protocol
 - Per la gestione della posta quando il programma di posta locale non risiede sul computer dell'utente ma su quello del suo *provider*
 - Gestisce l'accesso dell'utente alla casella di posta, fornendo funzioni di autenticazione utente e di gestione della casella di posta.

Posta via Web

- Vari portali di successo propongono il *mail server* integrato nel sito stesso, al quale accedere col protocollo http:
 - netscape.com, hotmail.com, libero.it sono solo alcuni esempi
- Vantaggio: accesso a un server *affidabile*
- Svantaggio: registrazione e pubblicità

Altri servizi via Web

- Tutti i browser implementano il protocollo HTTP, ma in genere sono multiprotocollo, sono in grado cioè di comunicare con altri tipi di server.
- In genere fungono anche da client per:
 - la posta (SMTP/POP/IMAP)
 - menu strumenti, posta in Internet Explorer
 - menu communicator, messenger in Netscape
 - il trasferimento di file (FTP)
 - es. <ftp://ftp.microsoft.com/>
 - le news (NNTP)

Software dalla rete

- **Freeware**: programmi distribuiti gratuitamente e utilizzabili da chiunque senza limitazioni
- **Shareware**: programmi distribuiti gratuitamente così che sia possibile provarli prima di acquistarli. Spesso hanno alcune limitazioni (temporali, funzionali) di utilizzo.