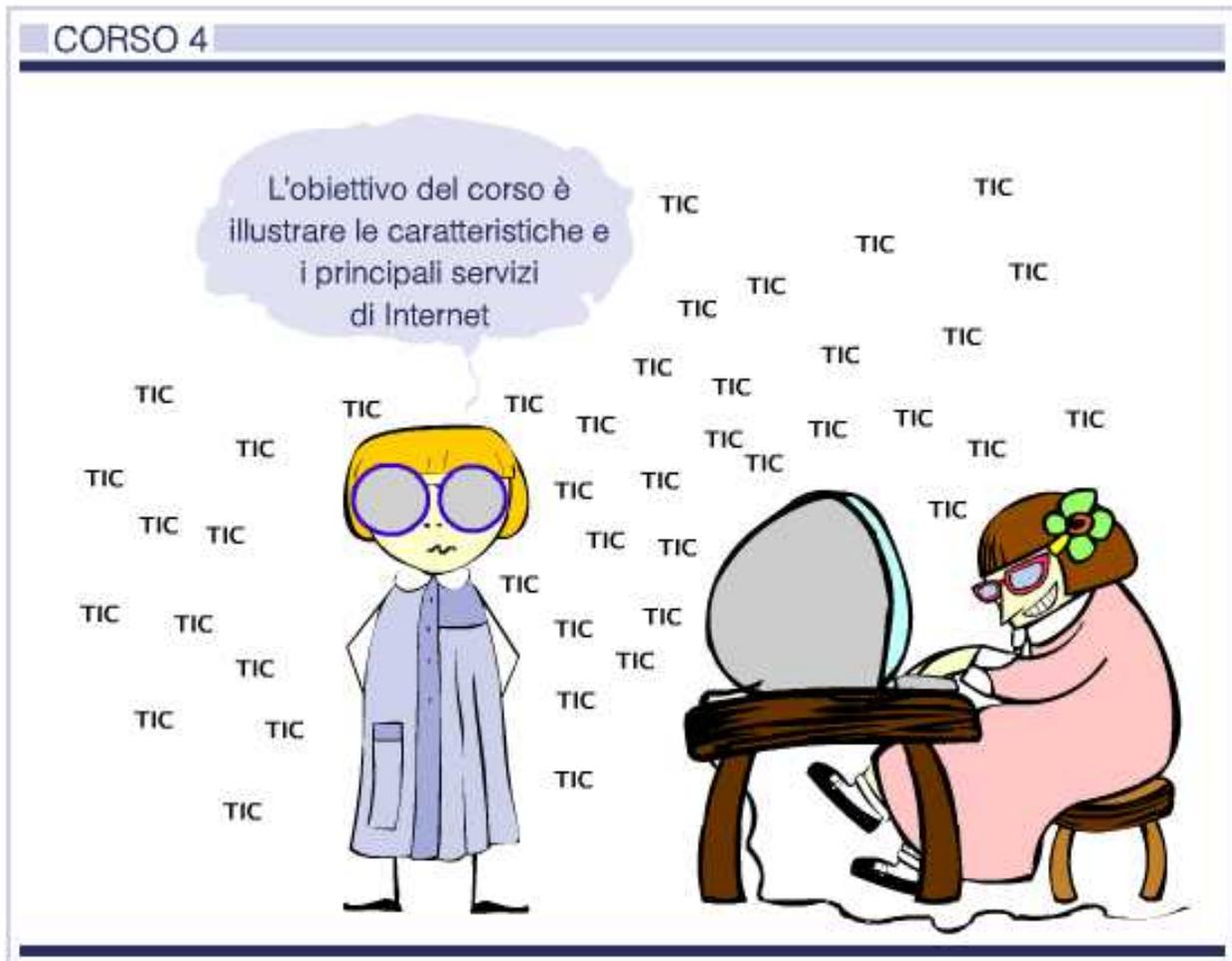


Introduzione alle Tecnologie Internet



Indice del Corso

1. Caratteristiche principali di Internet

- La storia
- Internet: cosa non è
- Internet: cos'è
- Gli standard
- La commutazione a pacchetto
- Protocollo TCP/IP
- Protocolli applicativi
- Indirizzo IP e hostname
- DNS

2. Teoria delle reti

- I nodi principali
- I canali fisici
- La linea ADSL
- Tipologie di reti
- Unità di misura
- La velocità di trasferimento dati

3. I domini

- Cosa sono i domini
- I Top Level Domain (TLD)
- I Top Level Domain generici
- I Top Level Domain nazionali
- Registration Authority (RA)
- I domini di secondo livello

4. Navigare in Internet

- Il browser
- Uniform Resource Locator (URL)
- Iper testo
- Impostazioni del browser: opzioni di Internet
- Bookmark o Preferiti

5. Servizi basati su Internet

- I motori di ricerca
- La posta elettronica
- Mailing list e newsgroup
- Netiquette e spamming

Modulo 1: Caratteristiche principali di Internet

L'obiettivo del modulo è fornire alcuni cenni storici e illustrare le principali caratteristiche di Internet e degli standard che ne hanno permesso la diffusione mondiale



Indice del Modulo

- La storia
- Internet: cosa non è
- Internet: cos'è
- Gli standard
- La commutazione a pacchetto
- Protocollo TCP/IP
- Protocolli applicativi
- Indirizzo IP e hostname
- DNS

La storia

Negli anni '60 il Ministero della Difesa statunitense crea un'agenzia per sviluppare una rete in grado di reggere ad un bombardamento nucleare: in questo modo la comunicazione tra luoghi diversi non viene interrotta. Dal progetto, che coinvolge centri di ricerca, università e aziende private, nasce ARPANET.

Grazie a importanti investimenti, questa rete cresce, aumenta la sua estensione e migliora le prestazioni in termini di qualità e velocità di trasmissione. Negli anni '80 diventa quello che oggi chiamiamo Internet.

Internet, acronimo di "INTERconnected NETwork", è l'interconnessione di milioni di reti e computer di tutto il mondo, che condividono informazioni tramite protocolli di comunicazione concordati.

Internet: cosa NON E'

Internet non è:

- una singola rete, ma un insieme di reti estese in tutto il mondo, la "Rete delle reti"
- governata da un gruppo, né da un ente, né da un'azienda
- gestita in modo centralizzato perché tutte le singole reti sono autonome
- sinonimo di World Wide Web*.

* **WWW**, letteralmente World Wide Web, è l'interfaccia di Internet che utilizza il linguaggio HTML decodificato da programmi (browser) per visualizzare documenti ipertestuali e multimediali.

Internet: COS'E'

Internet è:

- pubblica ma non gratuita
- un luogo dove si ottiene software gratuito
- un vettore della vostra posta elettronica (e-mail)
- ascoltare e vedere trasmissioni del vostro cantante preferito
- ascoltare e scaricare, legalmente, musica (FTP o Peer to Peer)
- comunicare con persone di tutto il mondo tramite chat, mailing list e newsgroup, forum
- avere un proprio diario personale visibile a chiunque (blog)
- un luogo per giocare on-line
- un luogo per compiere ricerche su qualsiasi cosa (motori di ricerca).

Gli standard

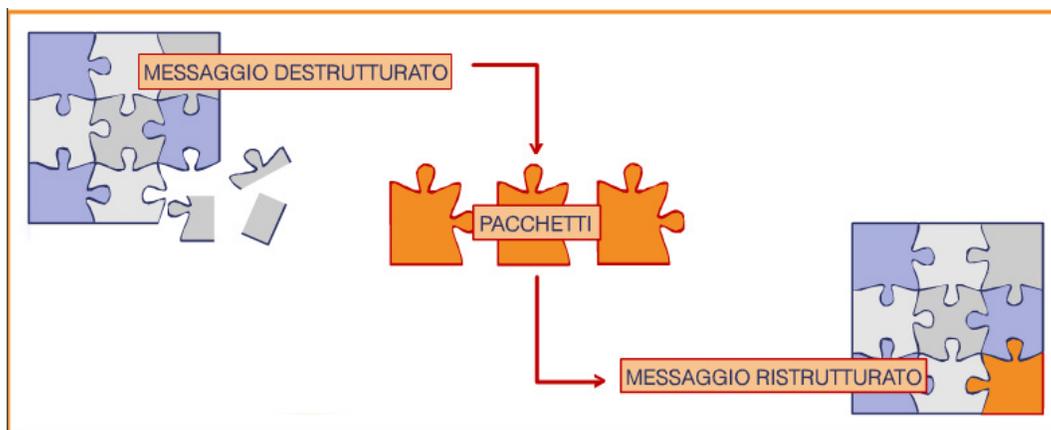
Internet è uno spazio fatto di migliaia di reti diverse (LAN o WAN) che usano differenti canali fisici di trasmissione (fibra ottica, doppino telefonico, satellite, onde radio, ...) e collegano diversi tipi di elaboratore (Personal Computer, mainframe,...). Nonostante questo, tali reti (network) riescono a comunicare tra loro poiché condividono un sistema di comunicazione (**commutazione a pacchetto**) e un linguaggio (i **protocolli TCP/IP**).

La Commutazione a pacchetto

I messaggi e le informazioni vengono divisi in pacchetti. Ogni pacchetto diventa un'entità a se stante che viaggia sulla rete in autonomia perchè possiede un'intestazione (**header**) con l'indirizzo di provenienza e di destinazione.

In viaggio i pacchetti possono dividersi e arrivare non con la giusta sequenza. Le informazioni che contengono bastano per ricostruire, all'arrivo, il messaggio originale corretto. Poiché i pacchetti che formano un messaggio seguono percorsi diversi, le possibilità che la rete rimanga congestionata diminuiscono (e aumenta la velocità di trasmissione) e qualunque sia lo stato della rete (guasti o interruzioni) il pacchetto può sempre arrivare a destinazione con una via alternativa.

I **Router**, dispositivi che instradano i dati da un punto all'altro della rete, sono in grado di leggere l'header del pacchetto per determinare il percorso di trasmissione migliore.



Protocollo TCP/IP

La suite di protocolli TCP/IP (letteralmente Transmission Control Protocol/Internet Protocol) è l'insieme delle regole che rendono possibile il dialogo tra più computer e la connessione ad Internet. TCP è il protocollo responsabile della suddivisione dei dati che vengono trasmessi in pacchetti, del rinvio di quelli non ricevuti, del riassetto dei dati.

Il protocollo IP assicura l'invio dei dati alla destinazione esatta: ogni elaboratore è identificato da un **indirizzo IP**.

L'IP è formato da una sequenza di 4 numeri separati da un punto. Ogni numero ha un valore compreso tra 0 e 255.

Ad esempio: 130.186.1.53

Protocolli applicativi

Sopra l'infrastruttura creata dalla rete TCP/IP vengono realizzati protocolli di alto livello che permettono il dialogo tra gli applicativi (ad esempio tra browser e web server).

Ai protocolli applicativi appartengono:

- **FTP** (File Transfer Protocol) protocollo standard per la trasmissione di file tra due sistemi in Internet. Si usa per scaricare i file da un computer remoto (download) e depositarli su un computer (upload).
- **HTTP** (Hyper Text Transfer Protocol) protocollo di trasmissione di documenti ipertestuali che ordina al server di visualizzare su un computer pagine web. È usato principalmente per scaricare pagine html.
- **HTTPS** (Hyper Text Transfer Protocol Security) protocollo utilizzato per lo scambio di informazioni riservate. È come HTTP ma con le comunicazioni browser – web server cifrate per impedire l'accesso alle informazioni.

Indirizzi IP e hostname

Per avere una comunicazione universale (ogni nodo della rete può comunicare con gli altri) serve un metodo con cui **identificare univocamente** ciascun *nodo* o *host*.

I dispositivi collegati ad Internet hanno due identificatori:

- **Indirizzo IP (numero di 32 bit)**: utilizzato per indirizzare ed instradare i pacchetti nella Rete
- **Hostname** (stringa alfanumerica): nome logico utilizzato dalle persone.

Indirizzo IP

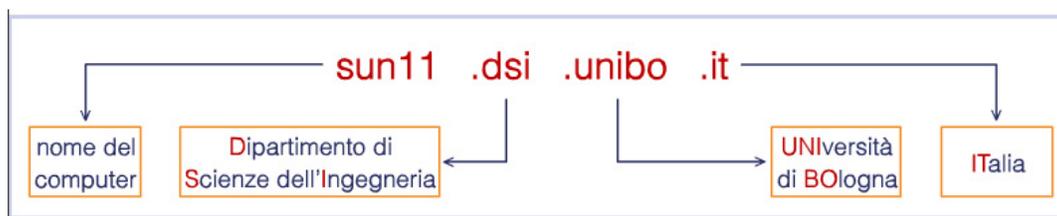
L'indirizzo IP è formato da 32 bit → $2^{32} \cong 4,3$ miliardi di indirizzi diversi. È rappresentato da 4 campi, ciascuno dei quali è formato da 8 bit, separati da un punto (*dotted notation*), ad esempio **160.80.15.72**

Ogni indirizzo è formato dalla coppia *netid*, *hostid*: *netid* identifica la rete e *hostid* uno dei suoi host. Tutti i nodi di una rete hanno lo stesso *netid* (prefisso di rete). C'è un registro con la lista degli IP assegnati per evitare che due elaboratori connessi abbiano lo stesso indirizzo.

Gli Hostname

Considerato che serve all'utente, all'hostname si preferisce attribuire valori mnemonici, ad esempio il nome del computer e la posizione.

Esempio di hostname:



Domain Name System (DNS)

È il servizio di Internet che converte i nomi dei computer in indirizzi numerici e viceversa.

Ad esempio:

Hostname: **sun11.dsi.unibo.it**
Indirizzo IP: **134.34.16.69**

È un ambiente client/server. I client interrogano un server DNS per convertire il nome di un computer. Se il DNS non è in grado di rispondere, gira la richiesta ad altri server fino a quando non scopre l'informazione. Dopo risponde al client.

DNS è simile a una rubrica telefonica. Quando la risposta arriva, viaggia a ritroso sullo stesso percorso fino al client che ha posto la domanda

Modulo 2: Teoria delle Reti



Indice del Modulo

- I nodi principali
- I canali fisici
- La linea ADSL
- Tipologie di reti
- Unità di misura
- La velocità di trasferimento dati.

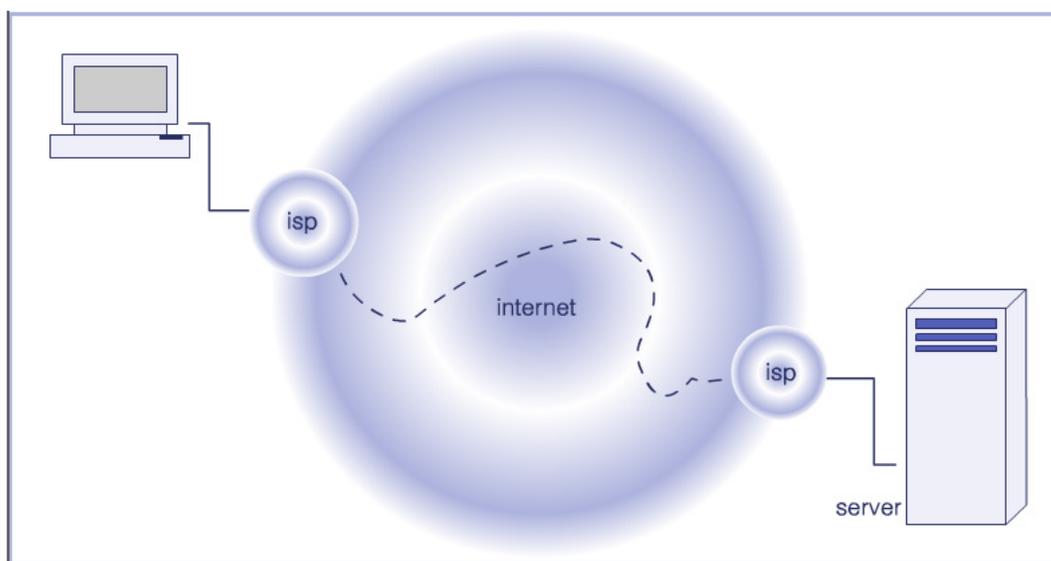
I nodi principali

I principali componenti di una rete sono:

- milioni di dispositivi connessi, detti nodi o host (ad es. PC, workstation, cellulari, server)
- canali di comunicazione: fibra ottica, radio, satellitari, rete elettrica, ecc.

Internet è composto da nodi con funzioni differenti. I più importanti sono:

- **Server:** è l'elaboratore che offre servizi, fornisce dati, ospita siti web, caselle di posta elettronica, ecc.
- **Client:** è l'elaboratore che, in un sistema client/server, inoltra le richieste dell'utente al server, ad esempio per visitare pagine web
- **Internet Service Provider (ISP):** è il soggetto che mette a disposizione dell'utente una rete (collegata a Internet) e dei servizi.



Approfondimento: Come scelgo un ISP?

■ Approfondimento > Come scelgo un ISP?

Per scegliere un ISP bisogna considerare:

- Prezzo
 - > Canone mensile, annuo, free
 - > Tariffe orarie o per megabyte trasferiti
- Supporto tecnico
 - > Telefonico
 - > Via e-mail
 - > Velocità di risoluzione dei problemi dell'utente
- Prestazioni
- Servizi aggiuntivi
 - > Indirizzi di posta elettronica "personalizzati", ad esempio mario@rossi.it, che se si cambia l'ISP rimangono invariati
 - > Spazio web (hosting) per pubblicare siti
 - > ecc.

I canali disponibili

I canali dati sono simili a tubi più o meno veloci a seconda del loro diametro. Pertanto:

- più grosso è il tubo, più dati passano
- basta una strozzatura per rallentare l'intero flusso.

I canali fisici di trasmissione dei dati possono essere:

▪ **Commutati**

Creano un circuito su richiesta e sono tariffati a distanza e tempo. Sono linee analogiche, come la linea del telefono combinata con un **modem**, e digitali, come la linea ISDN (Integrated Service Data Network) studiata per la trasmissione dei dati.

▪ **Permanenti** (o dedicati)

Linee digitali affittate per collegare due sedi in modo permanente. Sono tariffate a distanza e velocità, raramente a traffico, come i CDN (Circuiti Diretti Numerici).

Approfondimento: Il Modem

Approfondimento > Modem

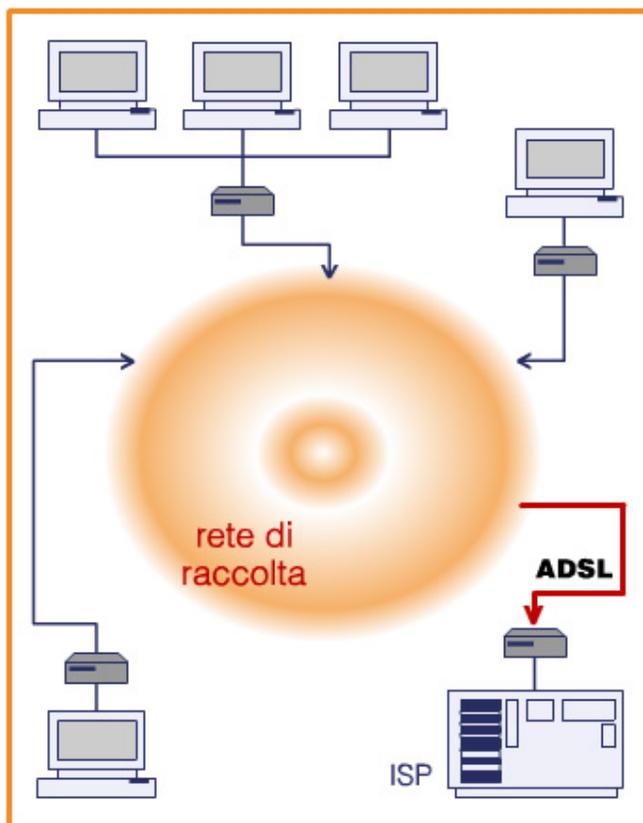
Il MODEM (MODulatore DEModulatore) è uno strumento che collega computer fisicamente distanti tramite le normali linee telefoniche.
Trasforma i segnali digitali (bit), propri degli elaboratori, in segnali analogici (suoni) tipici delle comunicazioni telefoniche.
Se due computer (chiamiamoli C1 e C2) devono scambiarsi informazioni ci vogliono due modem (rispettivamente M1 e M2), uno per ognuno, collegati alla linea telefonica.
Quando il computer C1 invia un messaggio, viene convertito (modulato) in analogico da M1, instradato sulla linea telefonica e riconvertito (demodulato) in digitale da M2, prima di essere disponibile su C2.
Se il mittente è C2 avviene il contrario.
I modem si classificano per velocità, in base ai bit al secondo (bps) che riescono ad inviare o ricevere.

La linea ADSL

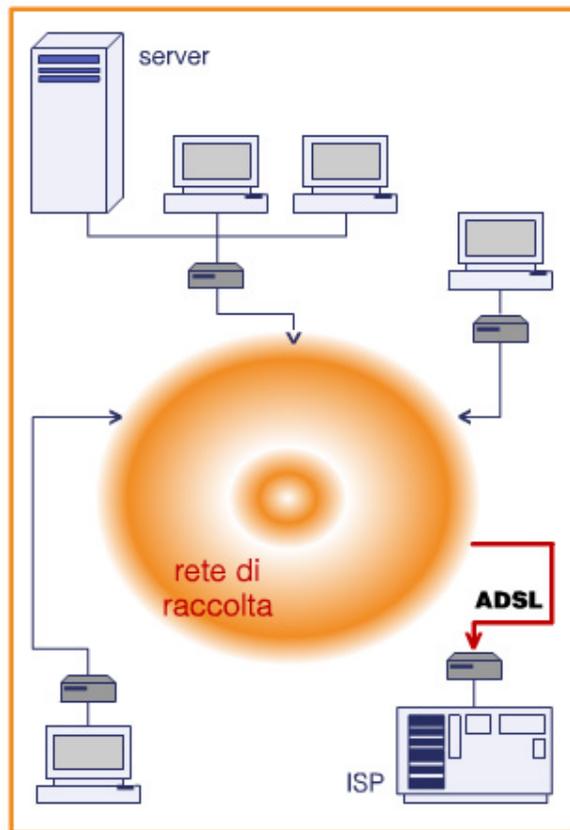
È un collegamento che raccoglie diversi utenti (connessioni) e li convoglia in un'unica destinazione, generalmente un ISP.

La linea ADSL viene frazionata in diversi "tubi virtuali" e i dati, scomposti in pacchetti, viaggiano parallelamente e contemporaneamente su tutti i canali, sfruttando al massimo la banda passante disponibile. L'intero blocco di informazioni viene ri-assemblato all'arrivo.

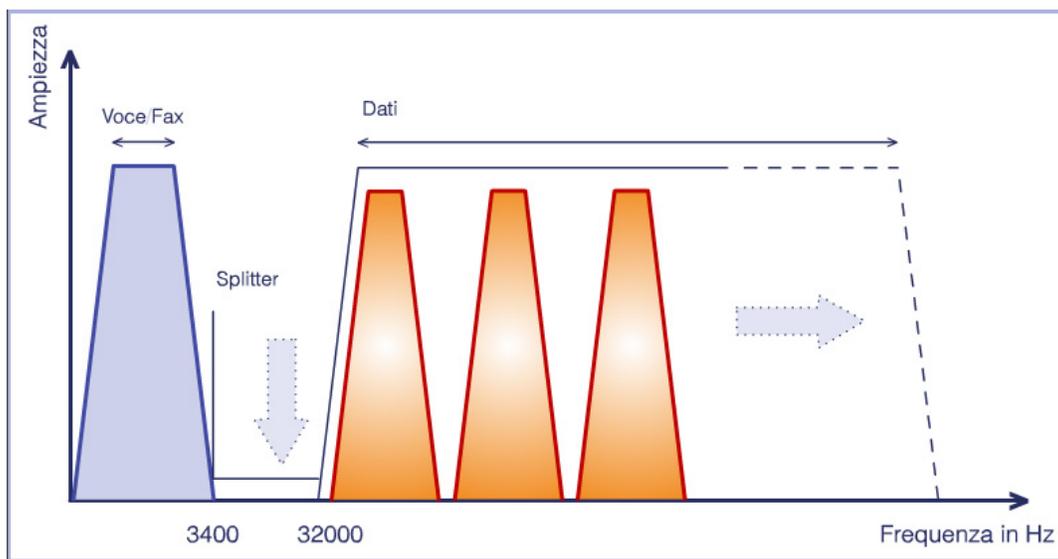
Se tutti gli utenti si collegano contemporaneamente con l'ADSL, la velocità di connessione è più lenta a causa dei vincoli del canale che collega la rete di raccolta degli utenti e il Provider.



La linea ADSL è asincrona, ovvero a velocità sbilanciata: i dati si trasmettono più rapidamente dal Provider all'utente che all'opposto. Si suppone infatti che il Client abbia bisogno di reperire dati, non di fornirne, quindi viene agevolato nella fase di ricezione più che in quella di richiesta. Per questo è sconsigliato collegare un Server dietro un ADSL, i tempi di risposta sarebbero troppo lunghi.



L'ADSL ha risolto il problema dell'*ultimo miglio*, cioè arriva nelle case degli utenti senza bisogno di nuovi cavi. Usa il doppino telefonico e nel cavo passano sia voce che dati. Si può parlare al telefono e navigare, insieme.



Tipologie di reti

■ Locali	---> Local Area Network (LAN)
■ Remote	---> Ambito urbano / Metropolitan Area Network (MAN)
	---> Ambito interurbano / Wide Area Network (WAN)

Reti locali

Sono installate in ambito locale, ad esempio in un'azienda perché non possono attraversare il suolo pubblico. Non sono a pagamento, quindi possono avere una velocità di connessione molto alta (da 10 a 100 Mb/sec fino a 1Gb/sec).

Reti remote

Sono fornite da gestori autorizzati dal governo (i soli che possono posare cavi sul suolo pubblico), che devono adottare i protocolli standard internazionali. Il costo della fornitura è calcolato in base a combinazioni di velocità, distanza, tempo di connessione, quantità di dati trasmissibili. Queste reti hanno una velocità medio bassa (da 64Kb a 2Mb al sec).

Le unità di misura

Per valutare le prestazioni delle reti bisogna definire come misurare la velocità di trasmissione dei dati. L'informazione elementare è gestita dal calcolatore con un dispositivo che assume solo 2 posizioni, per questo viene chiamata binaria (**bit**) e rappresentata con i due simboli **0** e **1**. Qualunque informazione all'interno del calcolatore è concepibile come una sequenza di bit del tipo 10010110100101001010... più o meno lunga, a seconda della sua dimensione.

Il bit è l'atomo dell'informatica.

Per rappresentare le lettere occorre una codifica che associa a una sequenza di bit (**b**) un carattere.

Ad esempio 00010001 rappresenta la A.

Per rappresentare tutto l'alfabeto il numero di combinazioni necessarie è di 8 bit: il **Byte (B)**.

- **b = bit** → atomo dell'informatica (0 o 1)
- **B = byte** (pari a 8 bit) → rappresenta un carattere

Per semplicità sono stati definiti dei moltiplicatori che rappresentano valori anche molto lunghi.

■ Tabella dei moltiplicatori	
K = kilo	-----> pari a 1024
M = mega	-----> pari a 1024K
G = giga	-----> pari a 1024M
T = tera	-----> pari a 1024G
P = peta	-----> pari a 1024T
E = exa	-----> pari a 1024P
Z = zetta	-----> pari a 1024E
Y = yotta	-----> pari a 1024Z

Ad esempio:

$$10 \text{ KB} = 1024 * 10 \text{ B} = 10240 \text{ B} = 1024 * 8 \text{ b} = 81920 \text{ b}$$

Di solito i **dati** si misurano in **byte**.

La **velocità** di trasferimento dei dati si misura in **bit al secondo**

- **Bps o b/s = bit al secondo**
- **Kb/s = Kilobit al secondo...**

La velocità di trasferimento dei dati

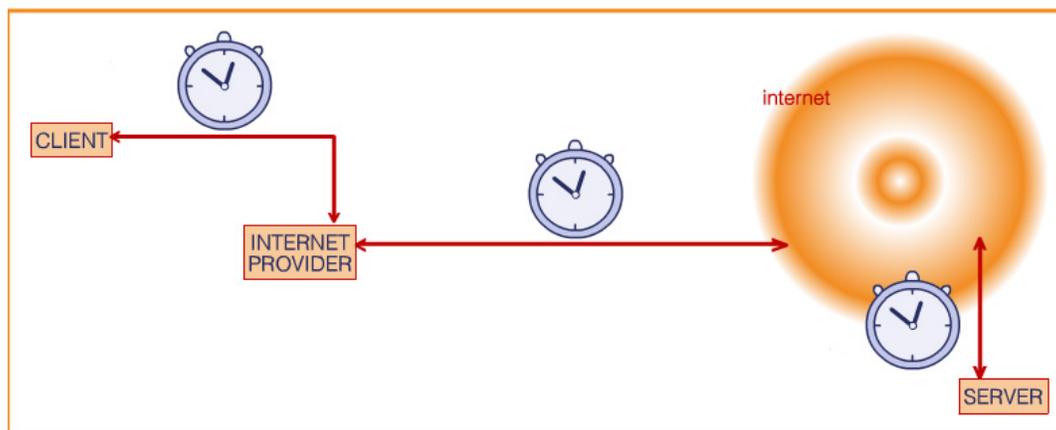
Illustrati i mezzi di misurazione, vediamo qual è la velocità di trasferimento a seconda della rete:

■ LAN	■ WAN	
da 100 Mb/s a 1 Gb/s	Linea Analogica	56Kb/s
	ISDN	64 Kb/s
	ADSL	da 640 Kb/s a 24 Mb/s
	CDN	da 64 Kb/s a 34 Mb/s

Perchè mi capita di navigare e scaricare da Internet lentamente?

Ora possiamo comprendere che la velocità di trasmissione dei dati dipende da diverse ragioni:

- dal tipo di connessione del computer client all'Internet Provider (analogica, digitale, ecc.)
- dall'intasamento delle linee che dall'Internet Provider vanno verso Internet
- dall'intasamento delle linee che dal Server vanno verso Internet.



Inoltre gli ISP non forniscono tutti gli stessi servizi. Ad esempio possono offrire **Garanzia di banda**, una velocità minima garantita, che è un'opzione in più nel contratto di fornitura della linea. E si paga.

Modulo 3: I Domini



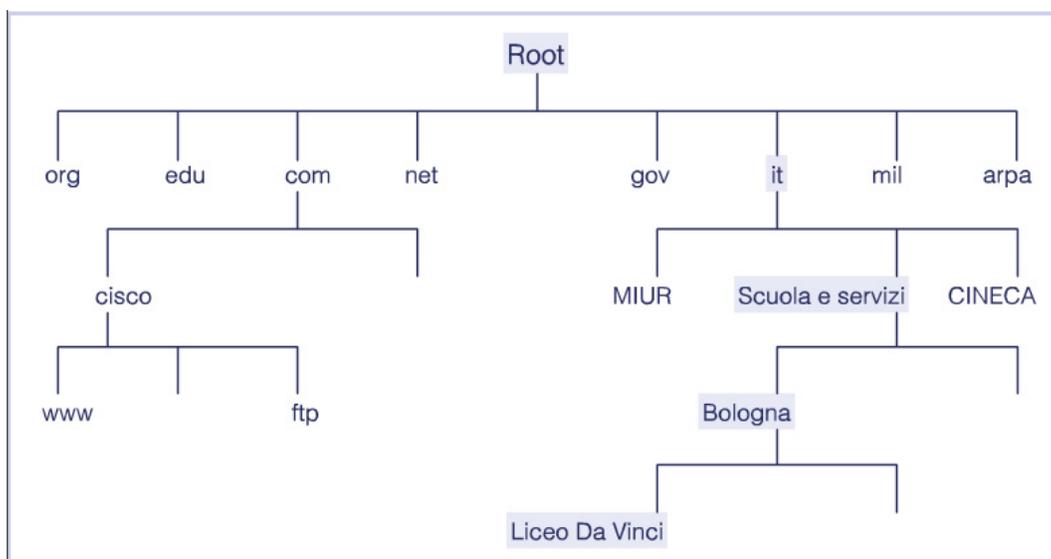
L'obiettivo del modulo è
illustrare,
in modo sintetico,
i diversi tipi di dominio e
le parti che lo compongono

Indice del Modulo

- Cosa sono i domini
- I Top Level Domain (TLD)
- I Top Level Domain generici
- I Top Level Domain nazionali
- Registration Authority (RA)
- I domini di secondo livello

Cosa sono i domini

In un sito il **dominio** è la parte significativa dell'indirizzo che segue la scritta WWW. I domini sono organizzati gerarchicamente in una struttura ad albero. Il vostro computer si trova nelle foglie.



I Top Level Domain (TLD)

Sono le parti finali dei nomi di dominio.

Si dividono in:

- **Generici (gTLD)** → Com org net Biz info
- **Nazionali (ccTLD)** → Legati ad una singola nazione: it fr de uk us eu.

La normativa attuale non pone limiti alla scelta del Top Level Domain. È intelligente (e utile in termini di marketing), adottare un TLD corretto: il dominio veicola un'identità in Rete, deve far capire subito chi sta comunicando.

I Top Level Domain generici

■ Tabella dei domini generici

Com	---> commercial
Net	---> originariamente per internet provider
Org	---> originariamente per organizzazioni no profit
Gov	---> agenzie governative USA
Mil	---> militari
Edu	---> istituzioni educative
Int	---> entità internazionali (poco usato)
Biz	---> ristretto al business
Info	---> utilizzabile da chiunque
Name	---> per registrazioni personali
Pro	---> professionisti quali avvocati, medici ecc...
Aero	---> chiunque abbia a che fare con il trasporto aereo
Museum	---> per i musei
Coop	---> per attività cooperative

I Top Level Domain nazionali

Ad ogni Nazione è stata assegnata una sigla di due caratteri:

■ Tabella dei domini nazionali

it	-----> Italia
fr	-----> Francia
de	-----> Germania
uk	-----> Gran Bretagna
no	-----> Norvegia
us	-----> Stati Uniti
eu	-----> Unione Europea

Registration Authority (RA)

È l'organismo che assegna i domini.

La RA cambia a seconda del dominio di primo livello scelto:

- Per il **.it** esiste la RA italiana
- Per quelli **generici** bisogna rivolgersi ad un registrar accreditato presso **InterNIC**.

IANA (Internet Assigned Numbers Authority) è l'autorità che sovrintende le altre, responsabile del coordinamento e della gestione del sistema DNS e dell'assegnazione degli indirizzi IP.

I domini di secondo livello

Chiunque possieda un dominio può creare sottodomini e delegarne la gestione ad altri.

Ad esempio: liceodavinci.bologna.scuolaeservizi.it

Modulo 4: Navigare in Internet



Indice del Modulo

- Il browser
- Uniform Resource Locator (URL)
- Ipertesto
- Impostazioni del browser: opzioni di Internet
- Bookmark o Preferiti

Il browser

È un programma che consente di visualizzare i contenuti ipertestuali che si trovano nel World Wide Web. Browser deriva dall'inglese "to browse" che significa visualizzare.

Tecnicamente un browser è un'applicazione client che utilizza il protocollo (ad esempio HTTP) per inoltrare le richieste dell'utente ad un web server e ottenere risposta. I più diffusi sono Microsoft Internet Explorer, Netscape Navigator, Mozilla e Opera.

I browser interpretano i **marcatori** (istruzioni **HTML** che definiscono la formattazione del testo) e le istruzioni di alcuni linguaggi di programmazione come:

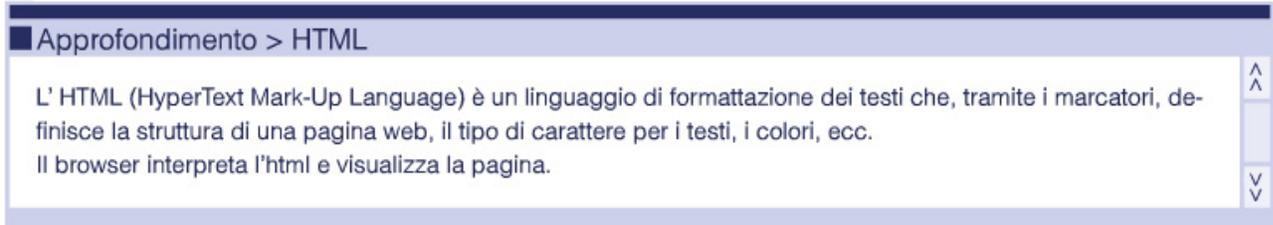
- **Javascript/VBScript** - usato soprattutto per migliorare l'interfaccia dei siti
- **Java** - con cui si creano veri e propri applicativi (sotto forma di applet) che vengono scaricati dal server ed eseguiti solamente su richiesta
- **Plug-in** - estensioni del browser con cui si possono visualizzare files altrimenti ingestibili (documenti PDF, animazioni flash, ecc.).

L'interfaccia del browser si compone di quattro aree principali:

- **Barra dei menù**: racchiude tutti i comandi utilizzabili
- **Barra degli strumenti**: con i pulsanti più utilizzati durante la navigazione e il campo in cui inserire la URL (o indirizzo internet) da visualizzare
- **Area di visualizzazione**: contiene il documento richiesto al server
- **Barra di stato**: indica quale operazione sta svolgendo il browser (se sta "chiamando una pagina", se ha completato l'operazione).



Approfondimento: HTML



Uniform Resource Locator (URL)

È una sintassi che identifica la posizione degli oggetti (testo, immagini, contenuti multimediali) su Internet.

Come: <http://www.cineca.it/directory/nomefile.html>

Nella forma più generale viene indicato:

metodo://user:password@nome_server/nome_directory/nome_oggetto

La URL si compone di vari elementi:

- **Metodo:** il protocollo da usare per trovare l'oggetto
- **Nome_server:** può essere letterale o numerico e indica il computer su cui si trova l'oggetto.
- **nome_dir/nome_oggetto:** è il nome, con il percorso completo, che identifica il file del documento nel server.

Esempio: www.cineca.it/area/risorsedicalcolo.htm

L'ipertesto

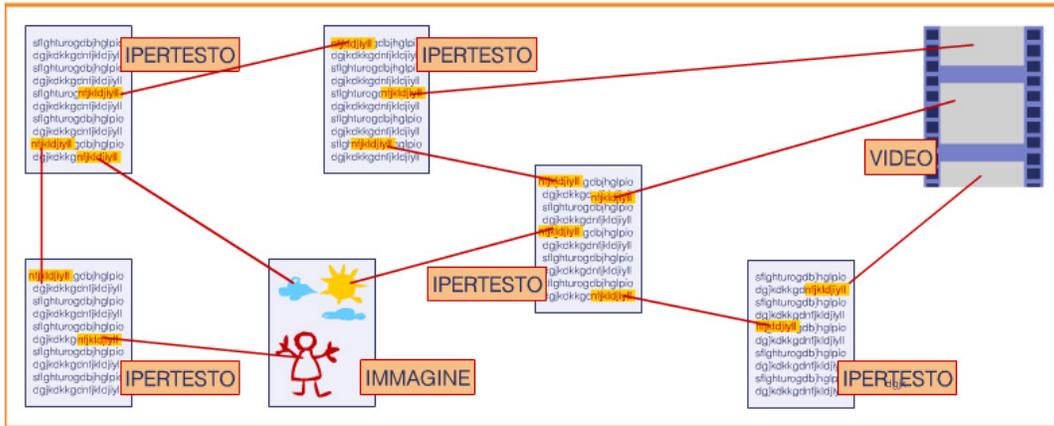
Le pagine visualizzate dal browser sono definite ipertesti.

Il termine deriva dall'inglese "hypertext": *hyper* viene dal greco e significa "sopra", *text* deriva dal latino *textum* (o *textus*) e significa "intreccio, trama".

Un ipertesto, dunque, è un testo "tridimensionale", o meglio un insieme di blocchi o frammenti collegati elettronicamente tramite una rete di interconnessioni semantiche non sequenziali. L'ipertesto è un nuovo sistema per organizzare e presentare le informazioni che va oltre il testo tradizionale, a cui accosta tutte le informazioni collegabili. Questi collegamenti si chiamano link.

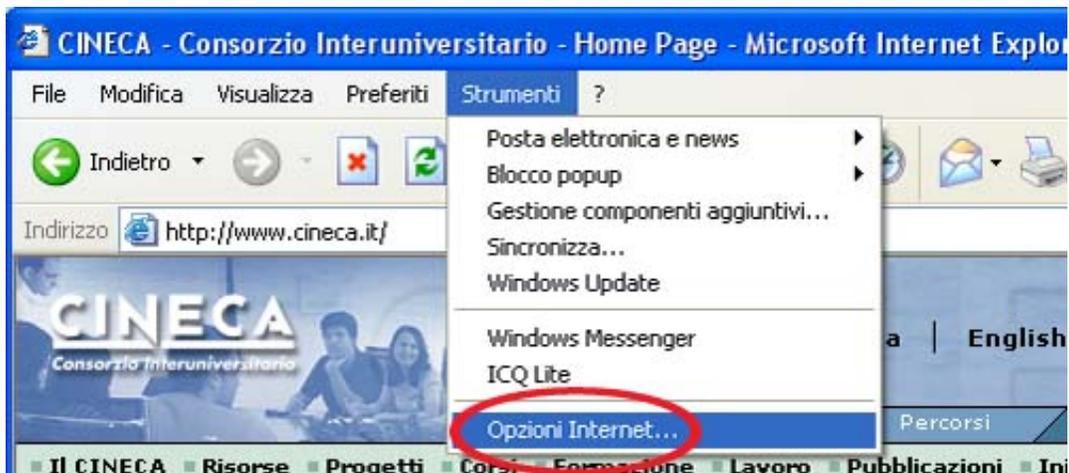
Visto il tipo di consultazione, gli ipertesti sono di solito di brevi e specifici. Rimandano gli approfondimenti ai link (anche collegandosi a pagine web realizzate da altri).

Le parole o frasi a cui sono applicati i link sono sottolineate e visualizzate con un colore diverso dal resto. Possono essere applicati anche a un'immagine o a una parte di essa (mappa sensibile) e, cliccandoci sopra, si ottiene l'accesso alla pagina o all'immagine, appunto, collegata.

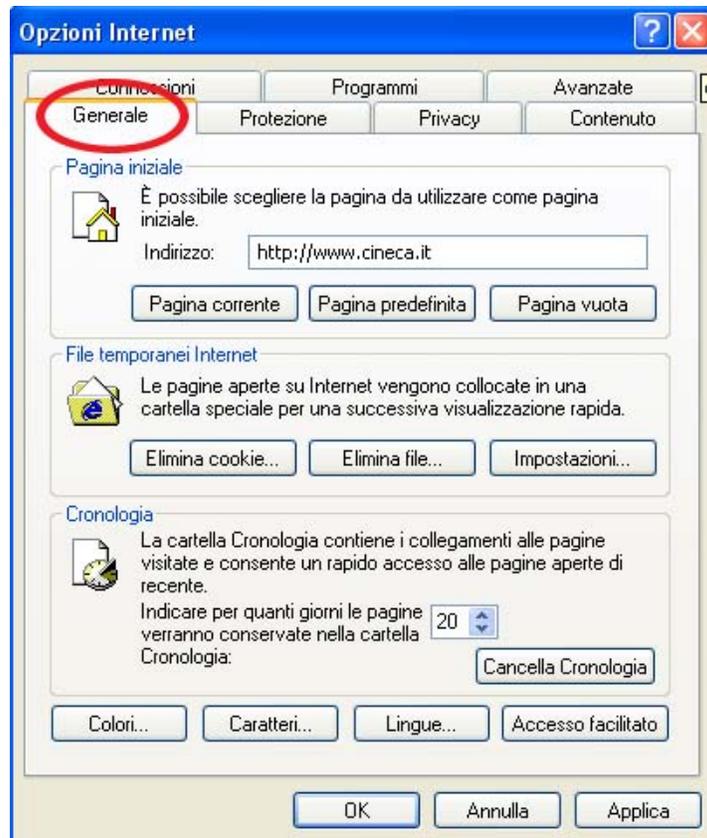


Alcune funzioni del browser

Di seguito illustriamo alcune funzionalità che consentono di migliorare e personalizzare la navigazione. Useremo, solo come esempio, l'interfaccia di Internet Explorer, sebbene le funzioni illustrate siano comuni a quasi tutti i browser.



Opzioni di Internet: Generali



Home page

È possibile inserire il link al sito che vogliamo visualizzare ogni volta che apriamo il browser.

File temporanei

Le pagine web visitate sono memorizzate dal computer nella **cache**, che è l'area di memoria consultata prima di qualunque richiesta ad Internet. Se il file richiesto dentro non c'è, il browser lo scarica dalla Rete, lo mostra e lo mette in cache per le prossime richieste.

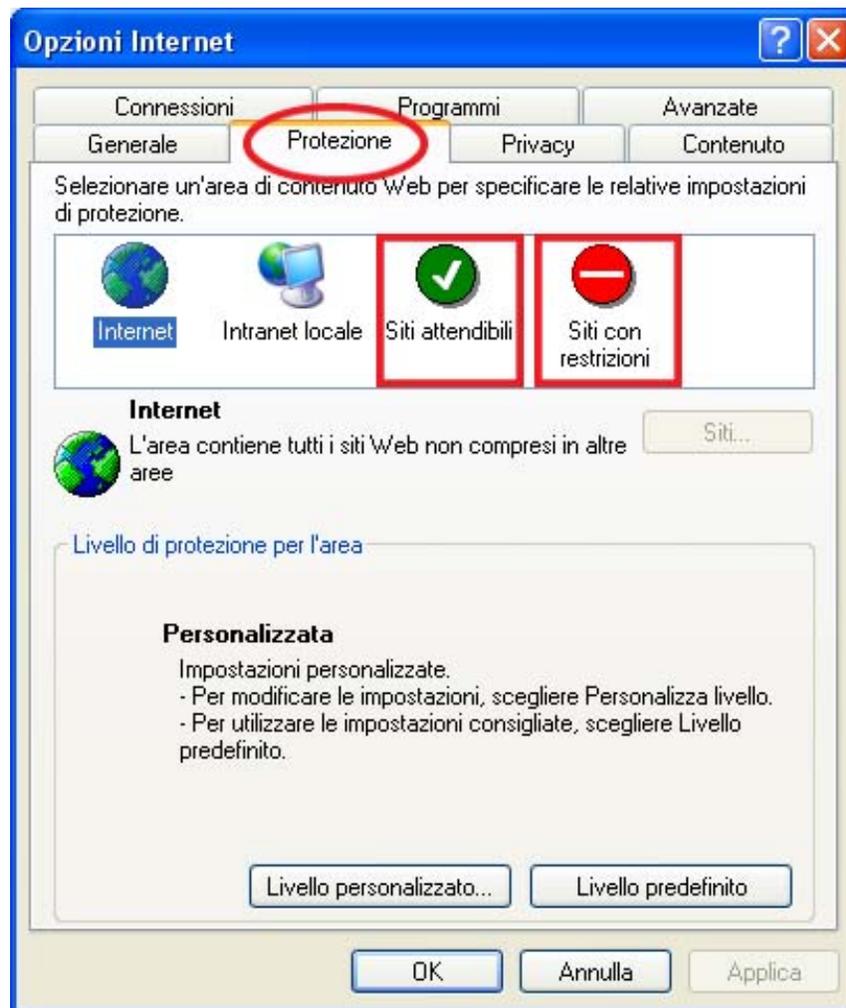
Un altro file temporaneo è il cookie.

I **cookies** (letteralmente "biscotto") sono file creati dal server Web che possono essere salvati quando l'utente visita un sito. Con il cookie il server manda tutte le informazioni che vengono rilette ogni volta si torni sul sito. Così un sito può adattarsi alle esigenze dell'utente e rendere più comoda la navigazione (ad esempio memorizzando la password di un utente, per evitargli ogni volta la digitazione).

Cronologia

Contiene gli indirizzi delle pagine visitate di recente. La barra di navigazione visualizza quelli memorizzati.

Opzioni di Internet – Protezione



Si può impostare il browser per comportarsi diversamente a seconda dei siti visitati, definendo:

- **siti attendibili**, quelli a cui si dà “fiducia”
- **siti con restrizioni**, quelli a cui si dà poca fiducia e di cui vogliamo limitare l’interazione (es. javascript).

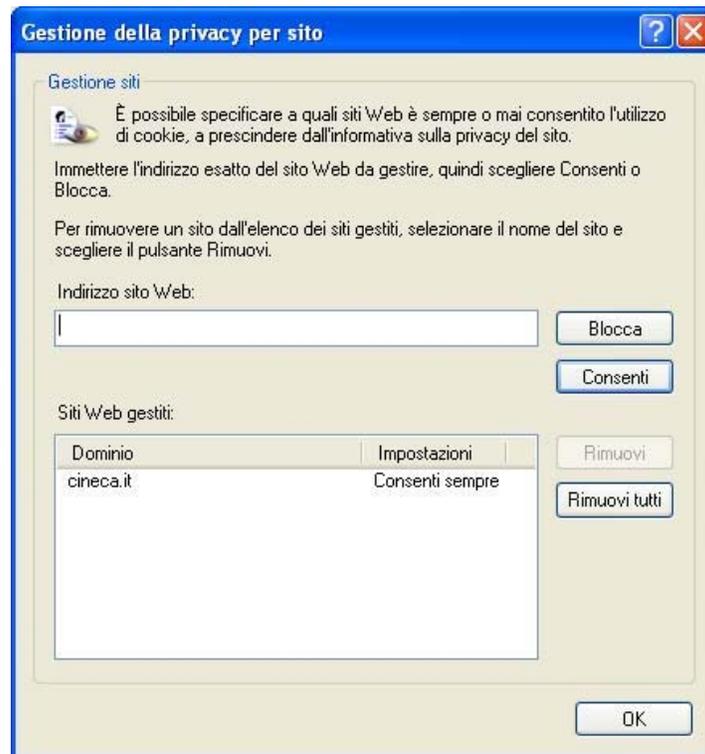
Opzioni di Internet: Privacy



Può succedere che aprendo una pagina venga visualizzata contemporaneamente una seconda finestra, magari sovrapponendosi a una parte di quella richiesta. Sono i cosiddetti **pop up** e di solito contengono messaggi pubblicitari.

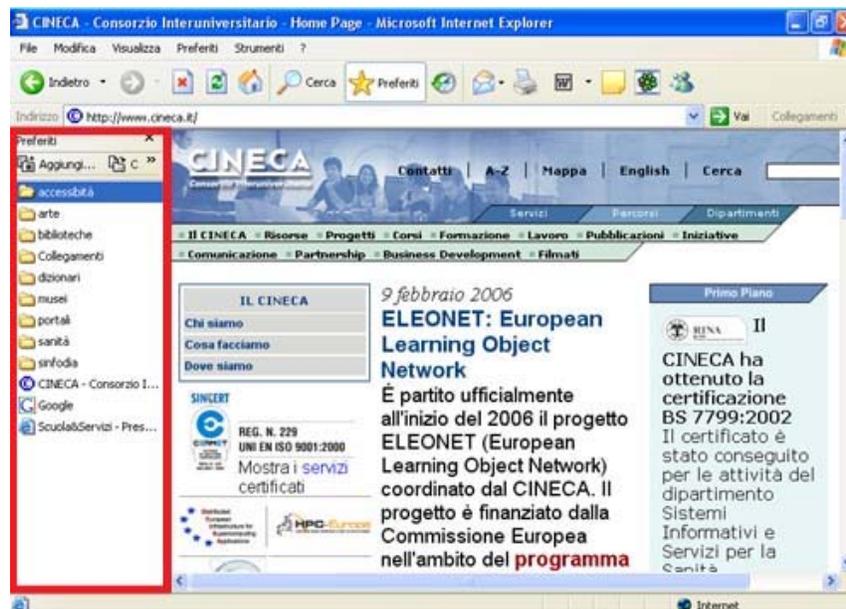
Se disturbano la lettura si può impostare il browser per bloccarli tutti.

Si può chiedere anche al browser di comportarsi diversamente a seconda della pagina: in particolare si possono definire i siti da cui accettare o bloccare i cookies.



Bookmark o Preferiti

È una raccolta di collegamenti URL a siti che abbiamo reputato di nostro interesse. Si consiglia di creare una struttura organica per facilitarne la ricerca.



Modulo 5: Servizi basati su Internet



L'obiettivo del modulo è illustrare i servizi Internet più usati. L'ultima lezione riguarda la *netiquette* e fornisce alcune regole di buon comportamento in Rete

Indice del Modulo

- I motori di ricerca
- La posta elettronica
- Mailing list e newsgroup
- Netiquette e spamming

I motori di ricerca

Internet è nato da un accorpamento progressivo di piccole reti. Questo significa disponibilità di una moltitudine di informazioni senza una gestione né un'organizzazione. Trovare l'informazione che cerchiamo è un problema che, con il passare del tempo e con l'aumento esponenziale della mole di dati, si è fatto sempre più pressante.

I motori di ricerca sono siti nei quali si possono trovare le informazioni. In teoria, a seconda del metodo di ricerca, si dividono in **motori classificati** e **motori spider**. In realtà oggi tendono ad operare in entrambi i modi, offrono servizi aggiuntivi, creano collaborazioni con altri motori, oppure implementano funzioni aggiuntive.

Dal momento che nessun motore fornisce una copertura completa dello scibile web e che una ricerca dipende in larga misura dal software implementato, si possono interrogare in successione più motori con le stesse parole chiave. In linea di principio dovrebbe essere efficace.

In realtà, otteniamo una ricerca poco efficiente e molto dispersiva: i risultati possono cambiare di molto oppure scoprire gli stessi link. Bisogna esaminare centinaia (migliaia) di possibilità, valutare i risultati utili e scartare i "doppioni".

Per questo è utile uno strumento di ricerca che possa fare tutto questo per noi: **il meta-motore**, un agente software di ricerca multipla, simultanea e comparata.

I Motori Classificati

È un motore che quando visita un sito lo inserisce in una categoria connessa al suo contenuto. La struttura "ad albero" delle categorie permette una suddivisione per tipologia, isolando quelli relativi alla ricerca. Alcuni motori classificati utilizzano macchine e software, altri fanno valutare ad un operatore umano il contenuto del sito.

I motori Spider

Sono motori che usano programmi automatici che visitano perennemente i siti e leggono tutte le pagine che incontrano a partire da una iniziale, seguendo ricorsivamente tutti i link in cui si sono imbattuti.

Ogni parola trovata viene inserita in un database in cui è indicato il sito fonte, il formato e la posizione relativa.

I motori spider (o robot) seguono degli algoritmi particolari. È opportuno sfruttare certe direttive se si vuole fare apparire il proprio sito fra i primi indicati da una ricerca sul motore spider.

I meta-motore

Come funziona un meta-motore e in cosa differisce da un normale motore di ricerca?

Un meta-motore trasmette la richiesta dell'utente sia ai motori di ricerca classici che ai loro database di pagine web. In pochi secondi restituisce una lista con gli esiti di tutti i motori, facendo una preziosa sintesi del risultato delle query. L'utente risparmia molto tempo.

I meta-motori non hanno un proprio database di pagine web, ma utilizzano quelli (e il software di ricerca) degli strumenti che interrogano.

Sono degli intermediari che, ricevuta la ricerca, la trasmettono, raccolgono i risultati e li organizzano criticamente in un unico resoconto.

Per lo stesso criterio di efficienza un meta-motore non può interrogare TUTTI i motori di ricerca esistenti (più di 900, si stima!): i tempi di risposta sarebbero inaccettabili. Perciò ognuno propone un bouquet di consultazione. L'esito è in ogni caso migliore di quanto si ottiene con un solo motore e rispecchia più fedelmente le effettive risorse disponibili in Rete.

La posta elettronica

È un mezzo asincrono che consente di inviare e ricevere messaggi da un computer collegato ad internet ad un altro collegato. Può spedire o ricevere posta chiunque possieda un mailbox (casella di posta elettronica).

Per utilizzare la posta elettronica dobbiamo configurare un programma con i parametri necessari.

I parametri minimi sono:

- **SMTP** server
- **POP/IMAP** server
- **username**
- **password**

SMTP sta per Simple Mail Transfer Protocol ed è il protocollo utilizzato per inviare i messaggi. **POP3** o **IMAP**, invece, servono per prelevare quelli ricevuti.

Invio e ricezione sono servizi distinti: un parametro errato in uno dei due non pregiudica il funzionamento dell'altro.

I server non mettono gli stessi limiti alla dimensione massima dei messaggi spediti e in ogni caso è fondamentale attenersi per il buon esito dell'invio. È utile sapere che, a causa delle limitazioni della posta elettronica, un allegato pesa circa il 30% in più di come risulta sul PC (sistema di encoding).

Anche la mail, sebbene con procedure diverse, va compilata con le stesse informazioni di una lettera tradizionale. Quasi tutti i nuovi programmi di gestione della posta elettronica permettono di inviare mail anche in formato html (oltre che in testo puro). In questo caso si può formattare il contenuto con dimensioni e colore, caratteri particolari, allineamenti ecc. Bisogna però verificare che il destinatario sia in grado di decodificare l'html.

La cassetta postale delle mail (mailbox) è identificata da un nome univoco e da una password così solo il proprietario può accedervi. Ogni mailbox può avere più alias, essere cioè raggiungibile con indirizzi diversi ed ha una dimensione massima in MB definita nel contratto con fornitore. Quando è piena non vengono più consegnati messaggi.

Mailing list e newsgroup

Con la posta elettronica si possono inviare messaggi personali ad altri utenti e riceverne. Quando questi messaggi dello stesso argomento vengono inviati a più utenti, un po' come una "circolare" abbiamo una **newsletter** o una **mailing list**.

Se i messaggi, invece di essere distribuiti, sono concentrati su uno o più computer (news server) che fungono da archivio, abbiamo i **newsgroup**. I diversi news server provvedono alla sincronizzazione dei messaggi dei gruppi per permettere la loro interazione.

I newsgroup sono gruppi di discussione nati su un tema definito. Per l'interazione si usa lo stesso software della mail.

I newsgroup italiani sono diverse centinaia e si riconoscono facilmente: hanno l'identificativo it nel nome. La gerarchia it.* è l'insieme dei gruppi di discussione in lingua italiana, creati dall'autunno del 1994.

Netiquette e spamming

In Internet regna un'anarchia ordinata: non esiste un'autorità centrale che decide cosa si può fare e cosa no, né organi di vigilanza. Per questo l'autodisciplina per usufruire della rete nel rispetto degli altri è a carico delle persone.

Negli anni si sono sviluppati fra gli utenti alcuni "principi di buon comportamento" conosciuti come **netiquette**.

Di seguito alcune regole da tenere presenti quando si utilizza la posta elettronica.

- È vietato lo **spamming** (invio di mail non richieste e di pubblicità)
- Rispettare il **copyright** sui materiali riprodotti
- Nel caso di **inoltro o re-invio** di un messaggio ricevuto, non modificarne il testo
- Non inviare mai **lettere a catena** via posta elettronica
- Fare attenzione quando s'inserisce l'indirizzo del ricevente
- Evitare di assumere **posizioni provocatorie**: è molto facile generare discussioni
- Verificare tutti gli indirizzi prima di iniziare discussioni lunghe o personali
- **Non divagare** rispetto all'argomento
- L'**oggetto** del messaggio dovrebbe riflettere il contenuto del testo
- Ci si assicuri della **grandezza** del testo inviato
- Ci si preoccupi di essere brevi senza essere eccessivamente concisi.