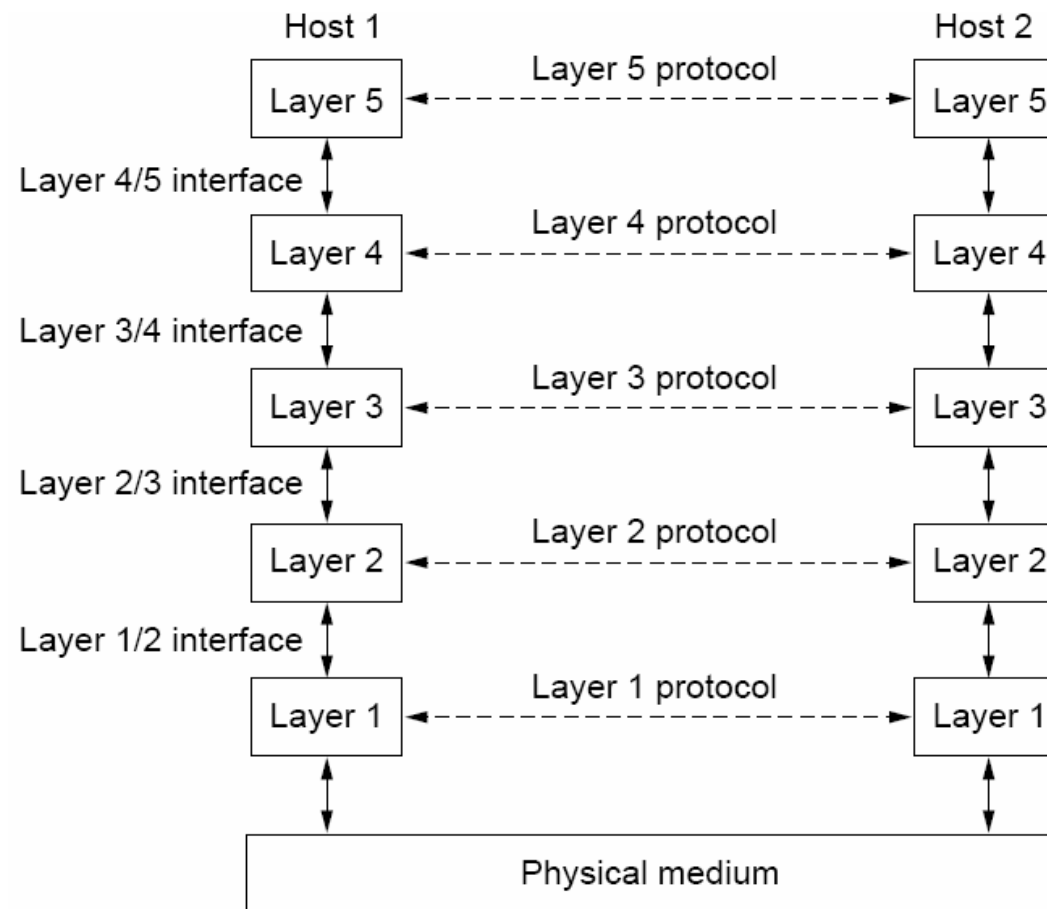


Reti di Calcolatori

Lezione 3

I livelli di una rete

Per ridurre la complessità di progetto, le reti sono in generale organizzate a livelli:



Il Formato dei Pacchetti

- Tutti pacchetti sono fatti in questo modo:



- Lo **header** contiene tutta l'informazione relativo al trasporto del pacchetto;
- La parte **data** contiene i dati che devono essere effettivamente trasportati;

Il Protocollo TCP/IP

- **TCP/IP** è un esempio di stack protocollare.
- Internet funziona con questo protocollo.



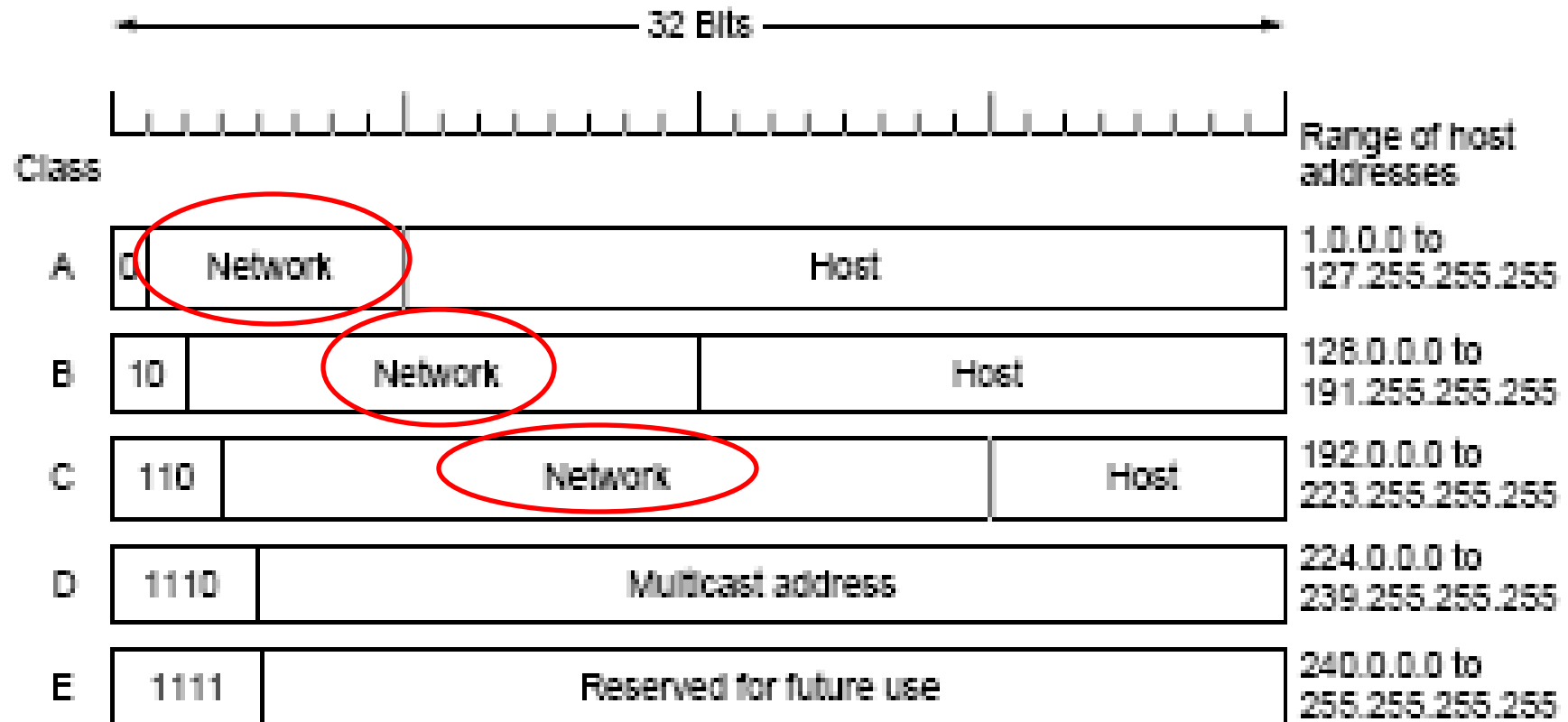
TCP/IP

La coppia di protocolli TCP/IP consente di mettere in collegamento due processi ovunque presenti su Internet:

- TCP divide il flusso di byte in pacchetti e ne garantisce un collegamento affidabile;
- IP immette i pacchetti nella rete ma non dà nessuna garanzia di affidabilità della rete, ma consente uno sfruttamento ottimale delle risorse;

Il Formato degli Indirizzi IP

Gli indirizzi IP sono del tipo: RETE-HOST;



Il Livello delle Applicazioni

II DNS

- Gli indirizzi IP sono in formato numerico: sono difficili da ricordare;
- Ricordare delle stringhe di testo è sicuramente molto più semplice;
- Il **Domain Name System (DNS)** consente di associare stringhe di testo a indirizzi IP;
- Il DNS si basa su un sistema di denominazione gerarchico;
- E' implementato mediante un base di dati distribuita;

Il Funzionamento del DNS

Vediamo come si usa il DNS:

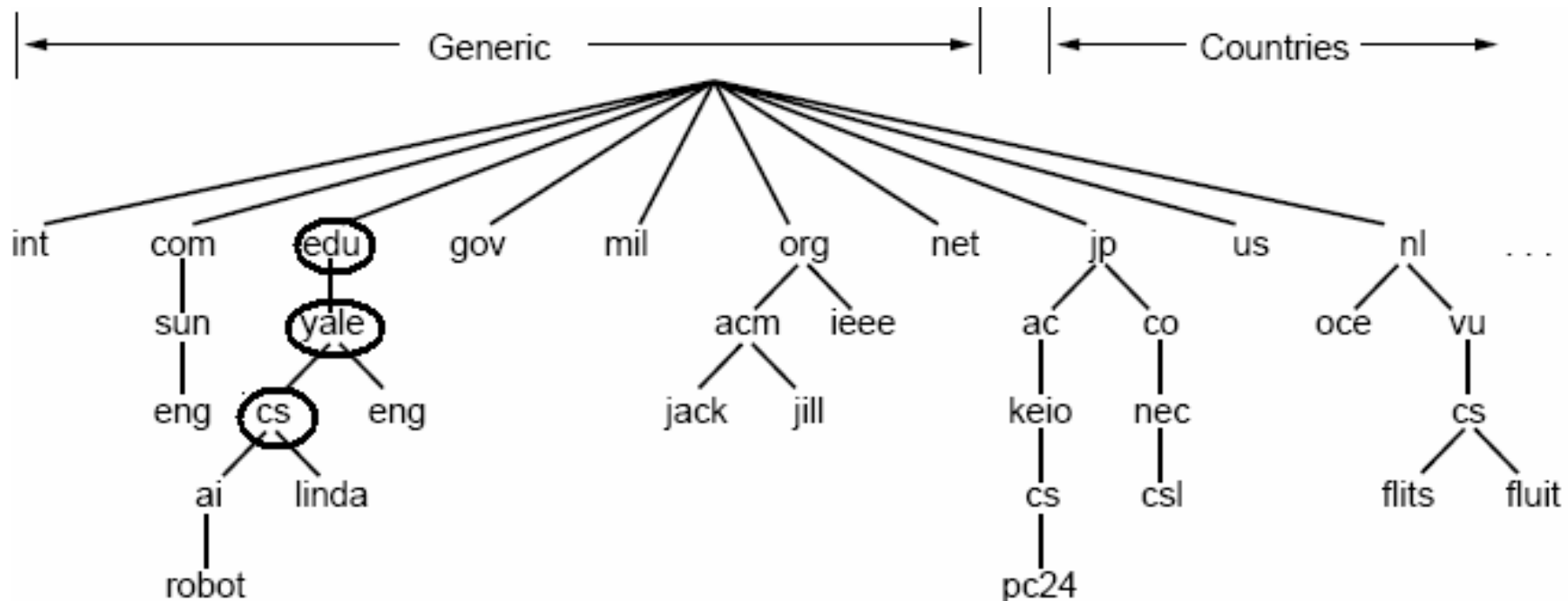
1. Quando un applicazione deve trasformare un nome in un indirizzo IP chiama una procedura risoltrice chiamata **resolver**;
2. Il resolver contatta un server DNS locale, il quale cerca nel data base distribuito;
3. Il server restituisce l'indirizzo IP al resolver;
4. Il resolver restituisce l'indirizzo all'applicazione;

Lo Spazio dei Nomi del DNS (1/2)

Il DNS si basa sul concetto di **dominio**:

- In Internet lo spazio dei nomi è suddiviso in domini radice, ogni dominio è a sua volta suddiviso in sottodomini, a loro volta frazionabili;
- Questa suddivisione può essere rappresentata con una struttura ad albero;

Lo Spazio dei Nomi del DNS (2/2)



Il nome di un dominio è composto dal cammino inverso dalla foglia fino alla radice.
Esempio: cs.yale.edu

Ogni dominio rappresenta uno sottospazio indipendente dei nomi ed il suo gestore è responsabile dei relativi sottodomini

DNS: La base di dati

- Il DNS si basa su una base di dati distribuita;
- Il DNS è basato su file di testo distribuiti in tutta la rete;
- Questi file sono organizzati in linee di testo;
- Ogni linea è divisa in colonne e contiene le informazioni relative ad un singolo dominio;

File DNS: un Esempio

NOME DOMINIO	INDIRIZZO IP	ALTRE INFORMAZIONI
unicas.it	193.205.60.2
Infn.it	193.206.84.198
Repubblica.it	213.92.16.191
Google.com	209.85.129.147

DNS: I Name Server

Il DNS è distribuito perché:

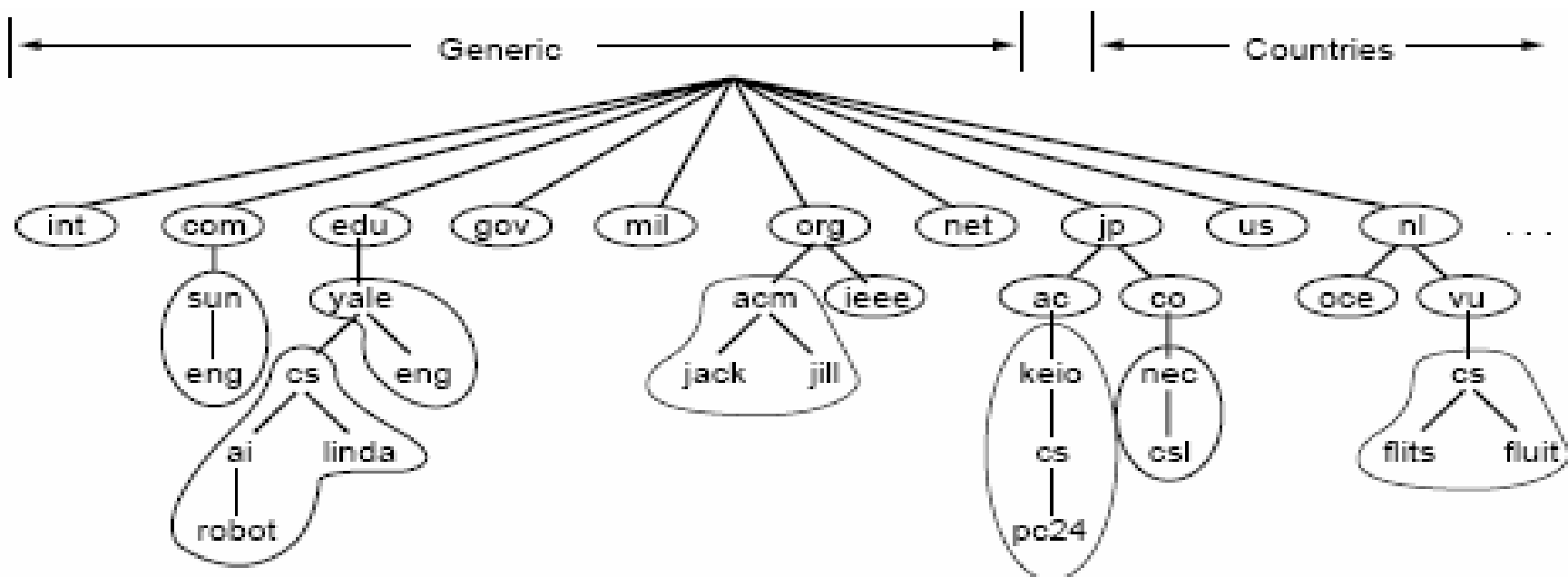
- Un unico server non riuscirebbe a rispondere a tutte le richieste;
- Il guasto dell'unico server metterebbe in crisi Internet;

Soluzione:

Il DNS è suddiviso in **zone** non sovrapposte;

DNS: La Suddivisione in Zone

Il DNS è suddiviso in zone non sovrapposte



Ad ogni zona sono associati uno o più name server

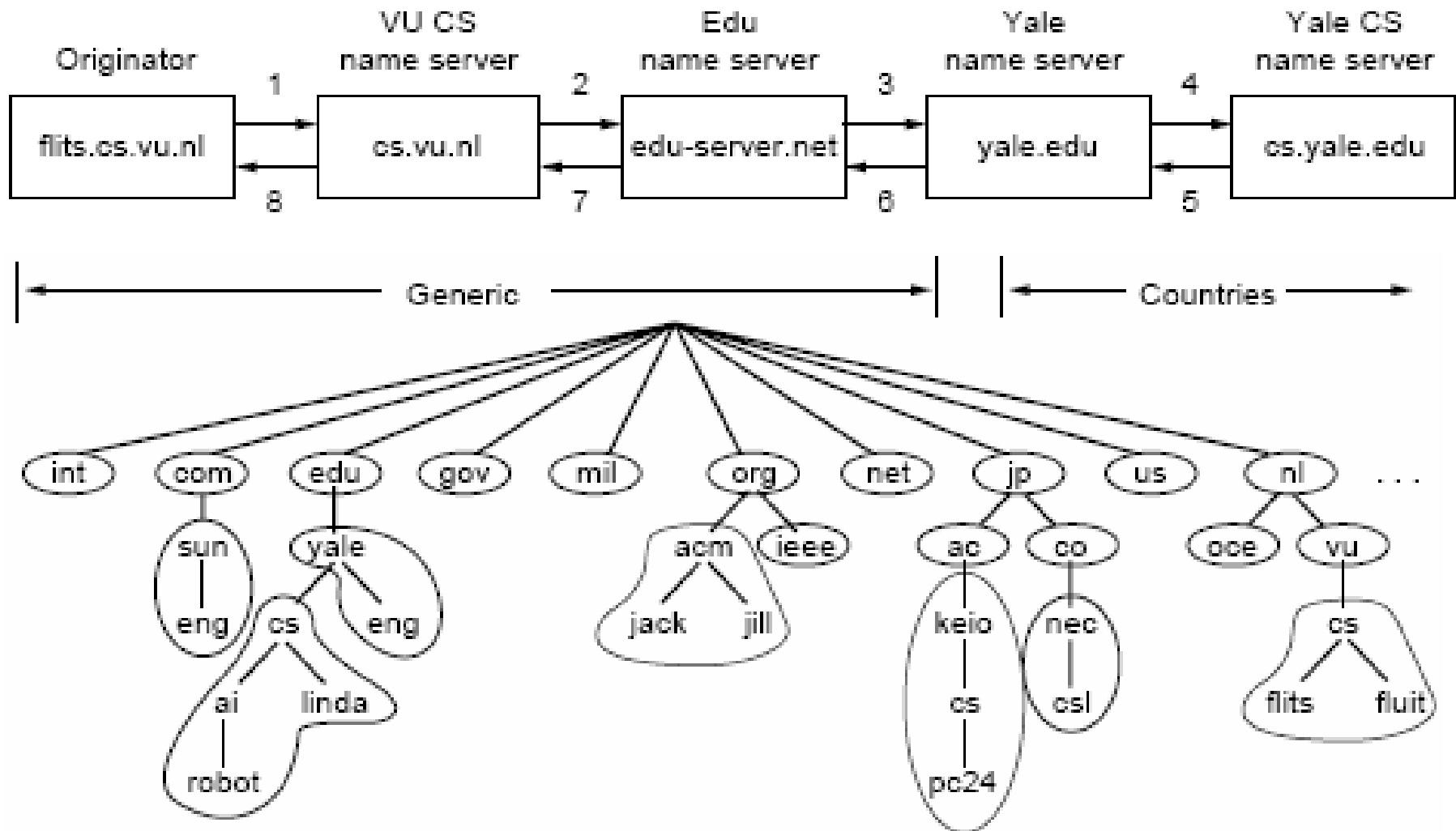
DNS: La risoluzione degli Indirizzi

Quando un resolver riceve una richiesta:

- Interroga un name server locale;
- Se il name server possiede l'informazione richiesta risponde, altrimenti interroga il name server di livello più alto;

DNS: un Esempio di Risoluzione

flits.cs.vu.nl richiede l'indirizzo di linda.cs.yale.edu



IL World Wide Web

- Il World Wide Web (WWW) è un architettura software che consente di accedere a documenti collegati tra loro e distribuiti sulle macchine presenti sull'intera Internet;
- Il web è nato al **CERN** di Ginevra nel 1989;
- Al CERN lavorano migliaia di ricercatori che devono condividere tantissimi documenti necessari alla realizzazione dei complessi esperimenti di Fisica Nucleare ad alta energia;
- Il suo inventore è: l'inglese Tim Berners-Lee;

WWW: il Lato Client

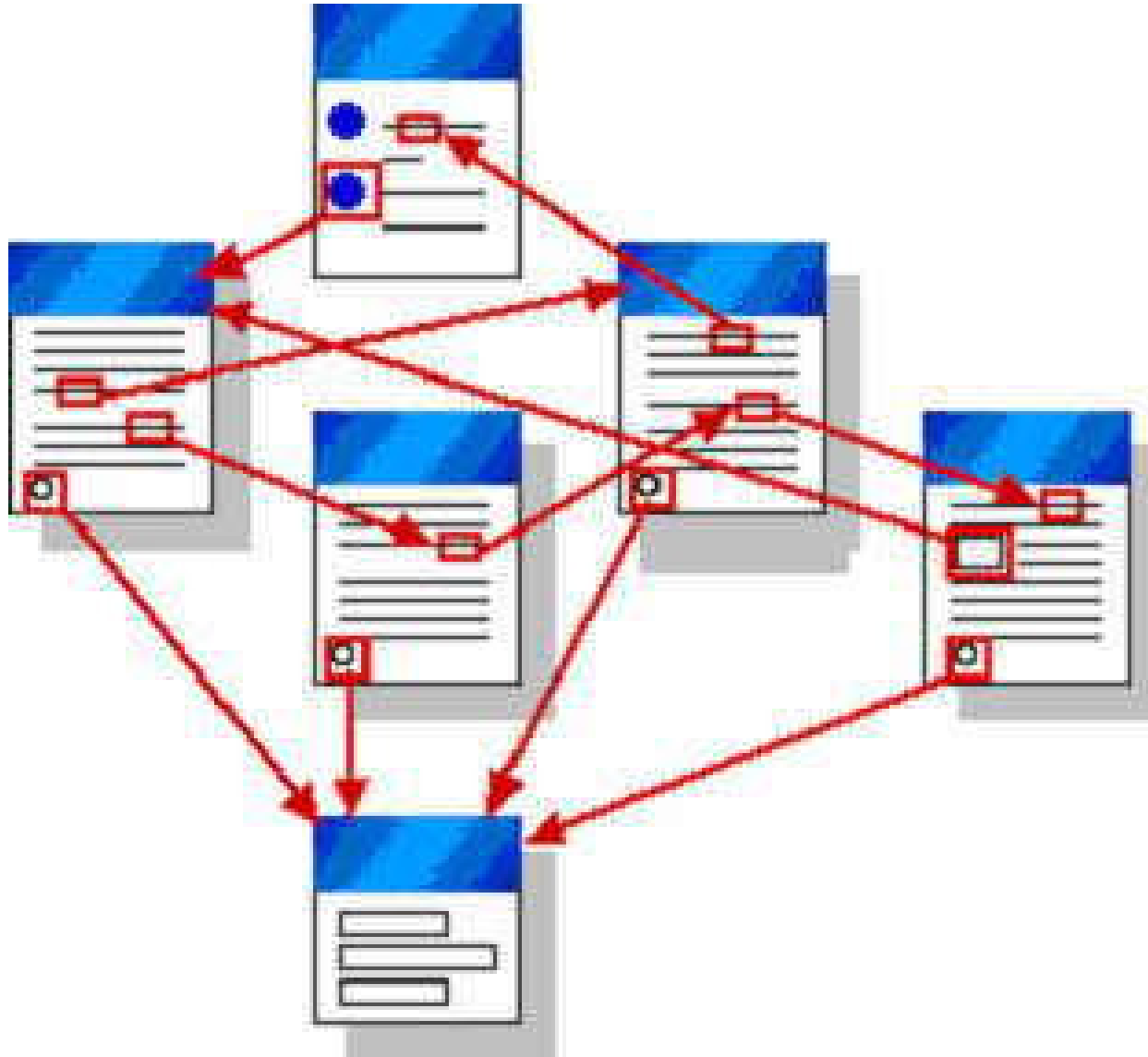
Dal punto di vista dell'utente il WWW è:

- Un enorme collezione di documenti, chiamati **pagine**, sparsi per il mondo (Internet);
- Le pagine sono dei file che si trovano su dei server;
- Un server è un normale PC sul quale è installato un programma che server;

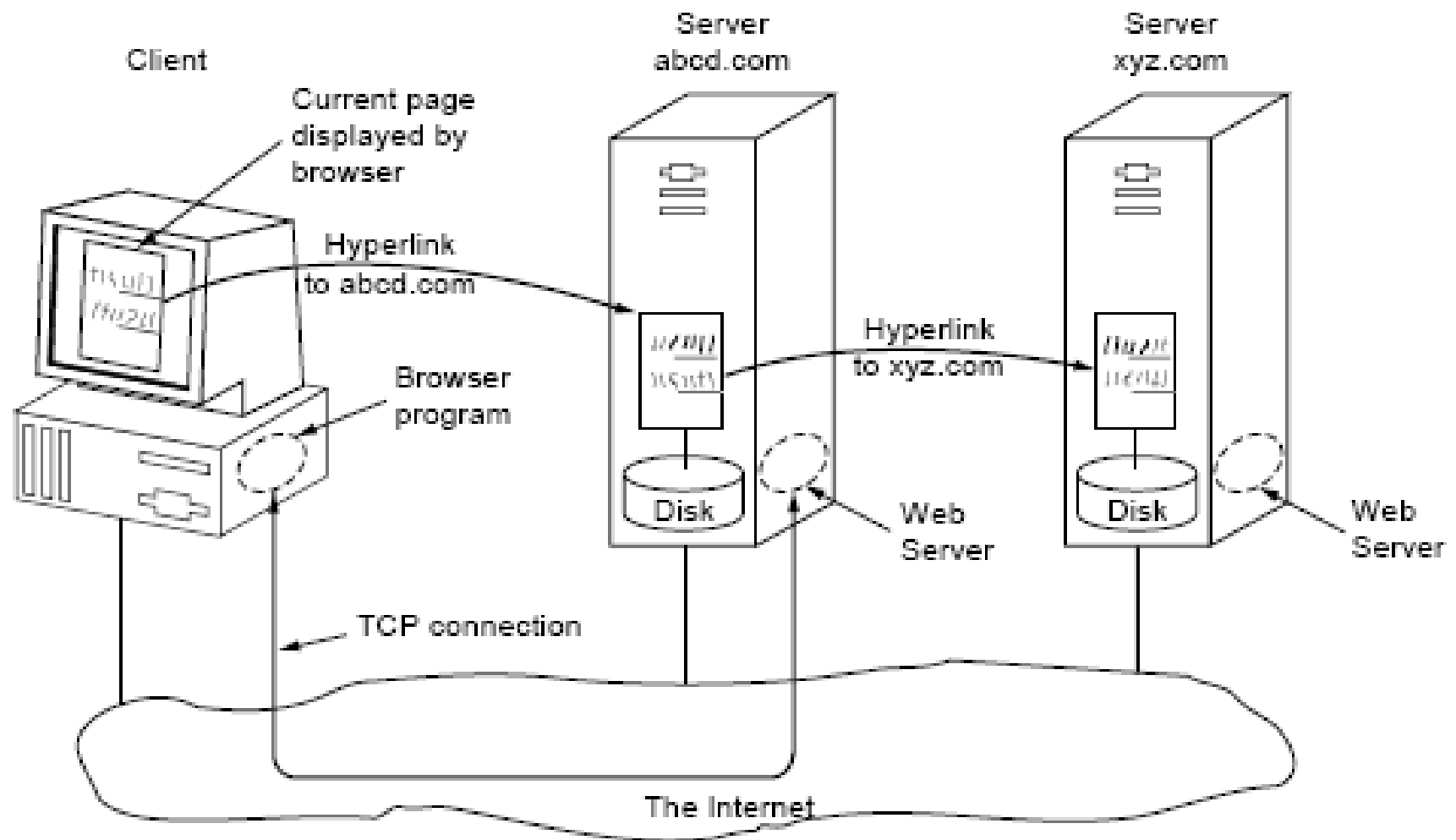
L'Iper testo

- Ogni pagina può contenere puntatori (link) ad altre pagine ovunque presenti nel mondo;
- Questa nozione di pagine che puntano ad altre pagine viene detta **ipertesto**;
- Le stringhe di testo che puntano ad altre pagine sono chiamate **iperpuntatori**

Il Concetto di Iper testo



WWW: la Realizzazione dell'Ipertesto



WWW: l'Indirizzamento

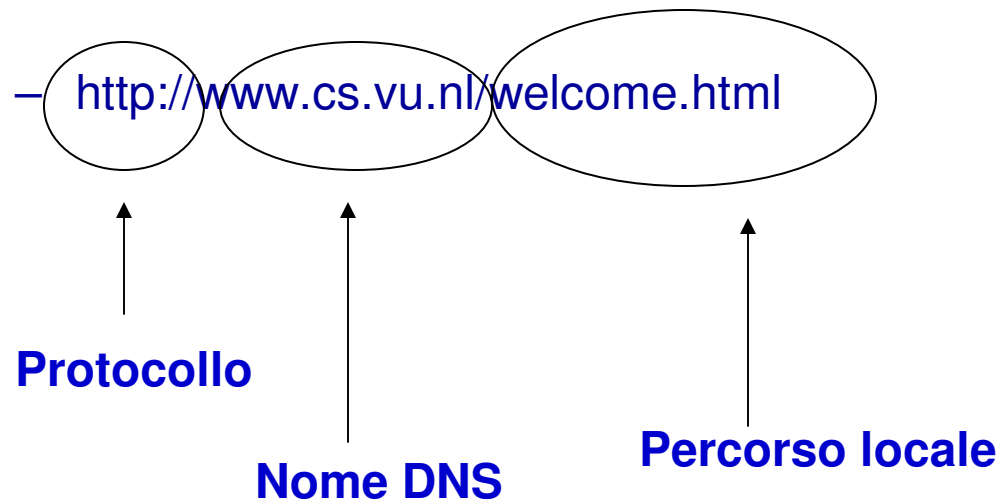
- Come sono implementati i puntatori? Ci sono tre problemi da risolvere:
 - Qual è la pagina richiesta?
 - Dove si trova la pagina?
 - Come si può accedere alla pagina?
- C'è la necessità di individuare in maniera univoca ogni pagina presente sul WWW;

Uniform Resource Locator

- La soluzione è quella di assegnare ad ogni pagina sul web un **Uniform Resource Locator (URL)**;
- L'URL è un nome di pagina univoco per tutta Internet;
- Un URL è formato da 3 parti:
 - Il protocollo;
 - Il nome DNS della macchina che ospita la pagina;
 - Un nome locale che indica la pagina specifica;

Un Esempio di URL

Un esempio di URL è il seguente;



NOTA: Gli URL vengono usati anche dagli altri protocolli;

Un Altro Esempio di URL

Se prendiamo l'URL

<http://webuser.unicas.it/fontanella>

- /fontanella è una cartella che si trova sul server HTTP che gestisce le pagine web dei docenti dell'Università di Cassino
- Nella cartella /fontanella sono memorizzate le pagine dell'utente fontanella;
- Sullo stesso server ci sono tante cartelle per quanti sono i docenti;

WWW: la Parte Server

- Un sito web, è costituito da:
 - un insieme di pagine (files),
 - Un processo server **in ascolto** sulla porta 80 della sua entità TCP;
- Il processo server attende richieste da processi client presenti su macchine della rete;
- Il protocollo utilizzato da client e server per scambiarsi informazioni è:

HTTP: Hyper Text Transfer Protocol

WWW: l'Apertura di una Pagina (1/3)

L'utente clicca su un oggetto (testo o immagine) che punta alla pagina il cui nome URL è

- `http://www.w3.org/hypertext/WWW/TheProject.html`

Cosa Succede?:

- Il browser chiede al DNS l'indirizzo `www.w3.org`;
- Il DNS risponde con `18.23.0.23`;
- Il browser si connette alla porta 80 di `18.23.0.23`;
- Il Browser invia il comando `GET/hypertext/WWW/TheProject.html`
- Il server HTTP invia il file `TheProject.htm`
- Si rilascia la connessione TCP;

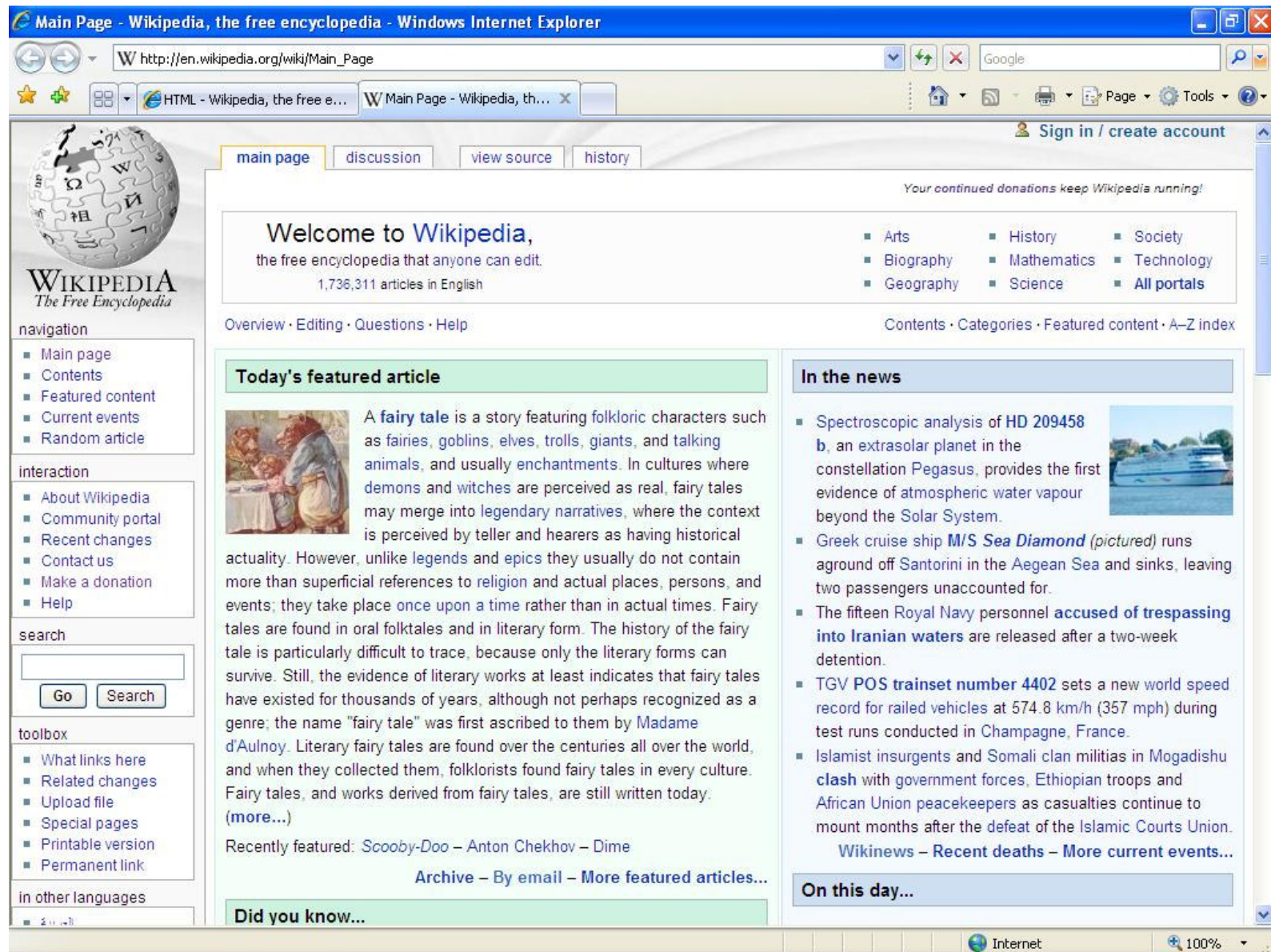
WWW: l'Apertura di una Pagina (2/3)

- Una volta ottenuta la pagina, il browser ne interpreta il contenuto del file TheProject.htm
- Il file contiene del codice HTML (Hyper Text Mark Up Language);
- Questo linguaggio è definito in maniera tale da consentire la corretta visualizzazione grafica dell'informazione contenuta nella pagina richiesta;
- Ha origine nelle tipografie dei giornali, dove i giornalisti segnavano le informazioni sulla formattazione a margine del testo da stampare;

Il Linguaggio HTML

- Una Pagina web è un file di testo;
- Il testo contenuto è speciale nel senso che esso è scritto usando il linguaggio HTML (Hyper Text Markup Language);
- HTML è linguaggio definito da uno standard aperto;
- HTML è definito in maniera tale da consentire la corretta visualizzazione grafica dell'informazione contenuta nella pagina richiesta;

Una Pagina Web



Il suo Codice HTML

```
Notepad++ - C:\Documents and Settings\francesco\Desktop\Main Page - Wikipedia, the free encyclopedia.htm
File Edit Search View Format Language Settings Macro Run TextFX Plugins Window ?
len30-den09-mut01.dat Risposta Aperta.txt Risposta Aperta.txt Main Page - Wikipedia, the free index.htm
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3c.org/TR/1999/REC-html401-19991224/loose.dtd">
2 <HTML lang=en dir=ltr xml:lang=en>
3 <?xmlns="http://www.w3.org/1999/xhtml"><HEAD><TITLE>Main Page - Wikipedia, the free encyclopedia</TITLE>
4 <META http-equiv=Content-Type content="text/html; charset=utf-8">
5 <META
6 content="Main Page,,1598,1742,1860,1943,1984,2007 Iranian seizure of Royal Navy personnel, Aegean Sea, African Union Mission to
7 Somalia, Anton Chekhov"
8 name=keywords><LINK href="/favicon.ico" rel="shortcut icon"><LINK
9 title="Wikipedia (English)" href="/w/opensearch_desc.php"
10 type=application/opensearchdescription+xml rel=search><LINK
11 href="http://www.gnu.org/copyleft/fdl.html" rel=copyright>
12 <STYLE type=text/css media=screen,projection>@import url( /skins-1.5/monobook/main.css?61 );
13 </STYLE>
14 <LINK media=print
15 href="http://en.wikipedia.org/skins-1.5/common/commonPrint.css?61" type=text/css
16 rel=stylesheet><LINK media=handheld
17 href="http://en.wikipedia.org/skins-1.5/monobook/handheld.css?61" type=text/css
18 rel=stylesheet><!--[if IE 5.5000]><style type="text/css">@import "/skins-1.5/monobook/IE50Fixes.css?61";</style><![endif]><!--[if IE 5.5000]><style type="text/css">@import
19 "/skins-1.5/monobook/IE55Fixes.css?61";</style><![endif]><!--[if IE 6]><style type="text/css">@import "/skins-1.5/monobook/IE60Fixes.css?61";</style><![endif]><!--[if IE 7]
20 <STYLE type=text/css>@import url( /skins-1.5/monobook/IE70Fixes.css?61 );
21 </STYLE>
22 <![endif]><!--[if IE 7]><script type="text/javascript" src="/skins-1.5/common/IEFixes.js?61"></script>
23 <meta http-equiv="imagetoolbar" content="no" /><![endif]>
24 <SCRIPT type=text/javascript>/*<![CDATA[*]
25 var skin = "monobook";
26 var stylepath = "/skins-1.5";
27 var wgArticlePath = "/wiki/$1";
28 var wgScriptPath = "/w";
29 var wgServer = "http://en.wikipedia.org";
30 var wgCanonicalNamespace = "";
31 var wgCanonicalSpecialPageName = false;
32 var wgNamespaceNumber = 0;
33 var wgPageName = "Main_Page";
34 var wgTitle = "Main Page";
35 var wgArticleId = "10282844";
36 var wgIsArticle = true;
37 var wgUserName = null;
```

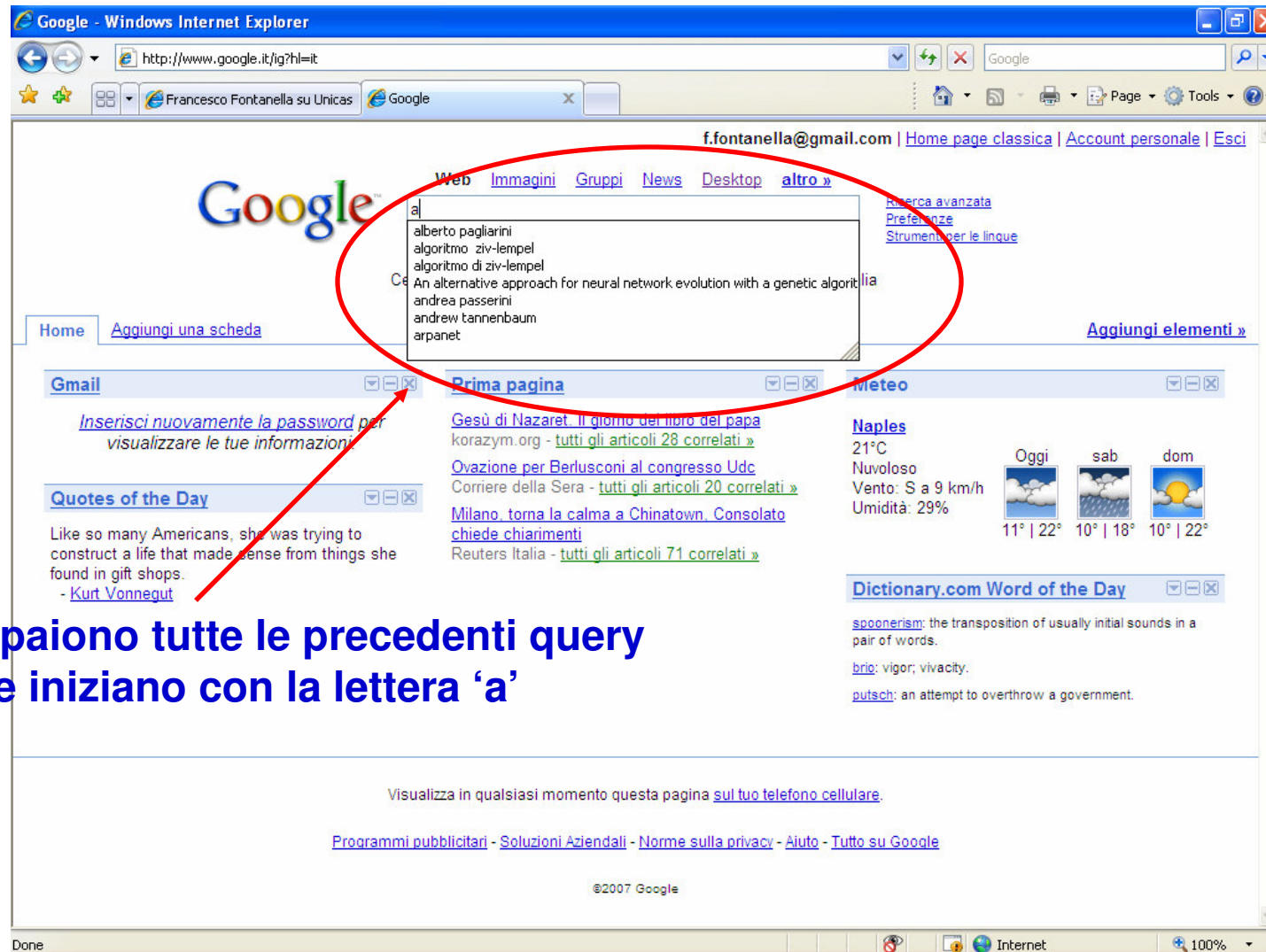

I Cookies (1/2)

- I **cookies** (letteralmente "biscottini") sono piccoli file di testo che i siti web utilizzano per immagazzinare informazioni sugli utenti;
- I cookie sono inviati dal sito web e memorizzati sul computer dell'utente che si connette;
- Sono re-inviati al sito web al momento delle visite successive;
- Le informazioni all'interno dei cookie sono spesso codificate e non comprensibili;
- Fanno parte dello standard HTTP;
- Contengono solo informazioni testuali e NON sono dei virus;

I Cookies (2/2)

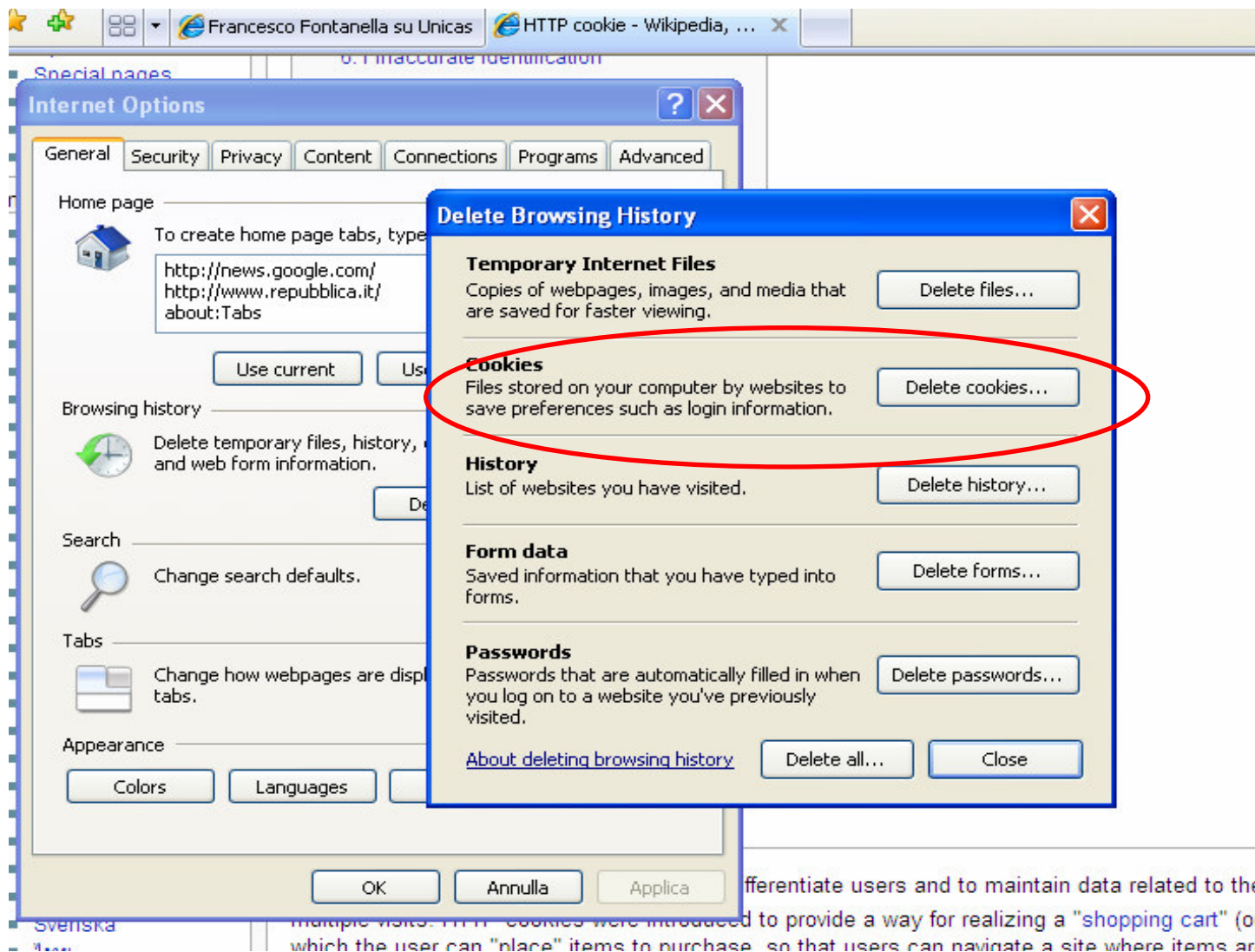
- Un cookie memorizza le azioni che un utente effettua nel visitare quel sito;
- Possono essere usati per:
 - Memorizzare i dati personali di un utente (nome utente, password, dati personali)
 - Personalizzare una pagina web;
 - Memorizzare il carrello della spesa nei siti di commercio elettronico;
- Hanno suscitato molti dubbi per quanto riguarda la **privacy**:
 - Accedendo ai cookies sul vostro computer è possibile avere delle informazioni su quali siti avete visitato e come li avete visitati;

Cookies: un Esempio

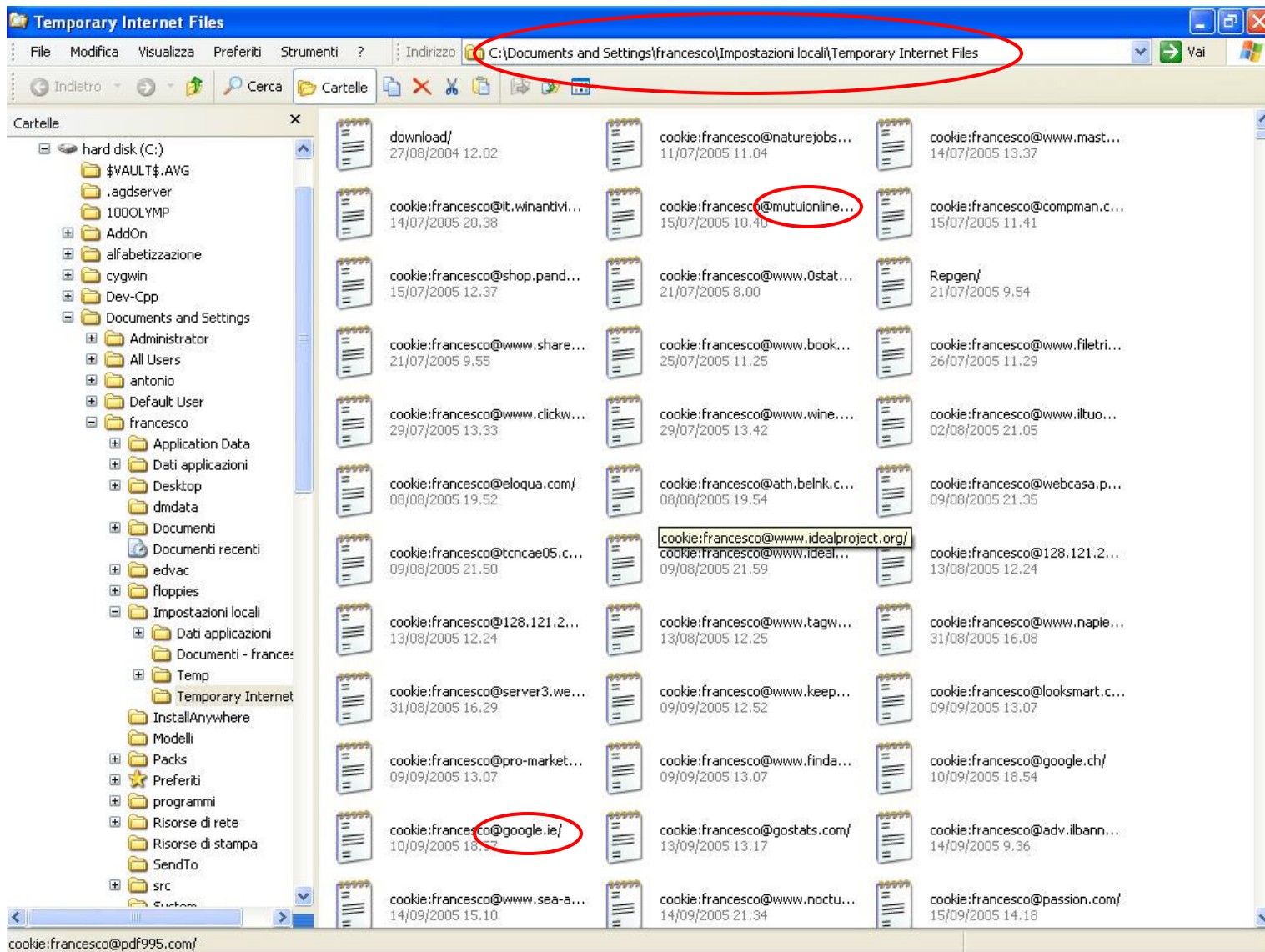


**Appaiono tutte le precedenti query
che iniziano con la lettera 'a'**

Cancellare i Cookies (Explorer 7)

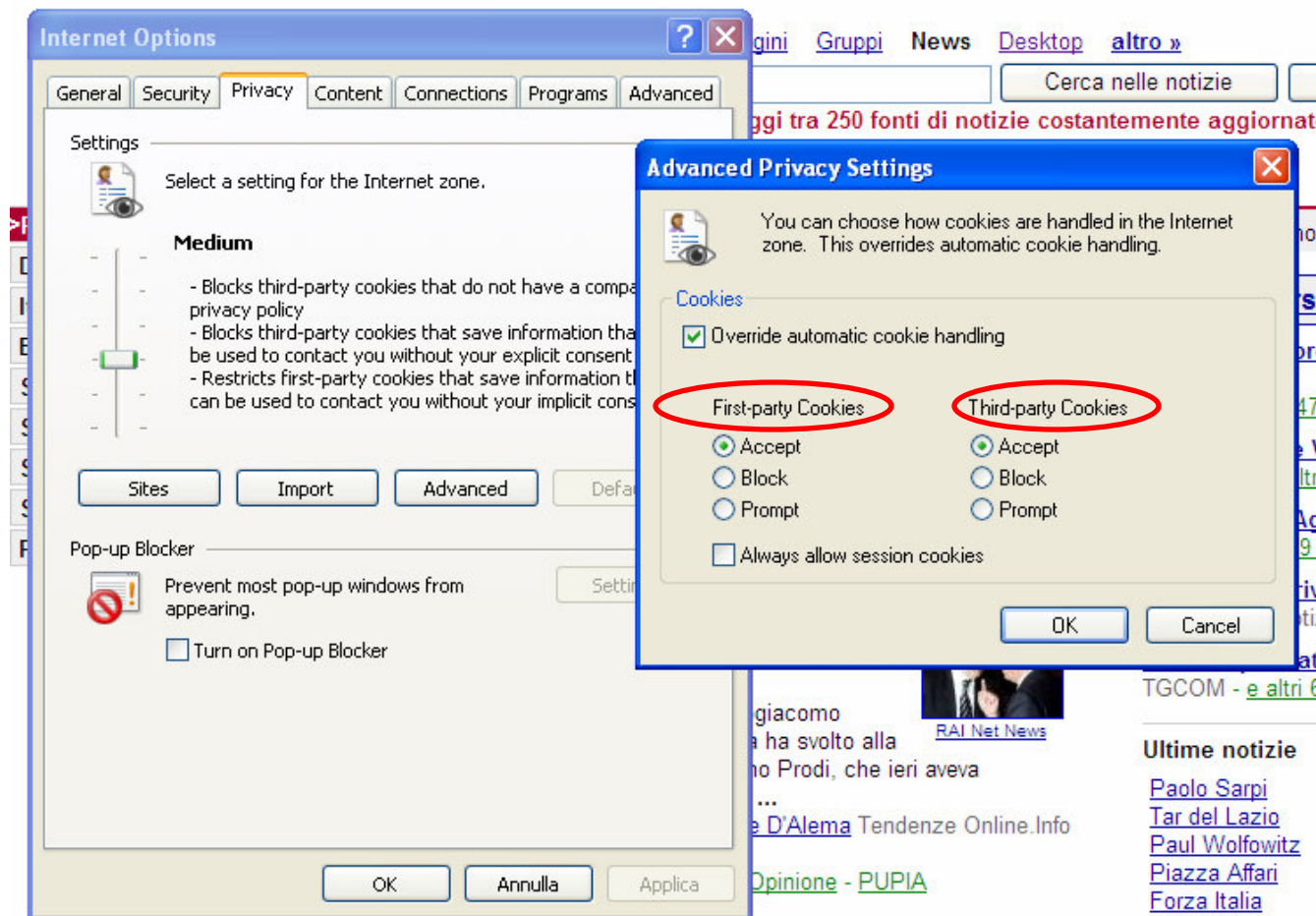


Cancellare i Cookies Manualmente



Monitorare i Cookies

- E' possibile impostare il rifiuto oppure la richiesta di conferma (prompt);



I Motori di Ricerca

- Il WEB contiene una quantità enorme di informazione.
- Senza uno strumento che ci aiuti a cercare (estrarre) le informazioni che vogliamo, tutta questa informazione è praticamente inutilizzabile;
- I motori di ricerca sono in grado di trovare l'informazione che stiamo cercando mediante **parole chiave** (key words);
- Il Lavoro fatto dai motori di ricerca è detto di Information Retrieval (Recupero dell'informazione);

I Motori di Ricerca

- Un **motore di ricerca** è un sistema **automatico** che analizza un insieme di dati spesso da lui stesso raccolti;
- Costruisce un **indice** dei contenuti disponibili classificandoli in base alla loro rilevanza.
- I motori di ricerca possono essere usati per cercare informazione in:
 - Il web;
 - Una rete Aziendale;
 - Il proprio computer;

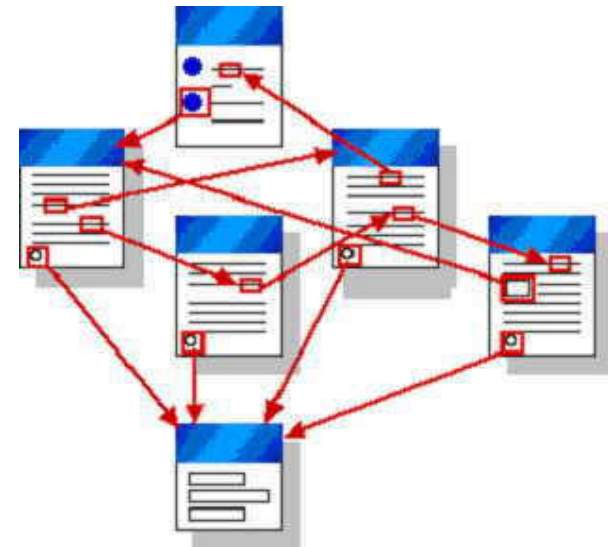
Come Funziona un Motore di Ricerca

Il lavoro dei motori di ricerca si divide principalmente in tre fasi:

- **Analisi;**
- **Catalogazione** del materiale ottenuto;
- **Risposta** alle richieste dell'utente;

L'Analisi

- Per analizzare il web i motori di ricerca utilizzano dei programmi detti **crawler** (o **spider** o **robot**);
- Un crawler visita un certo numero di URL contenuti in un elenco e segue i successivi link che si trovano all'interno dei documenti analizzati e che non fanno ancora parte dell'elenco;



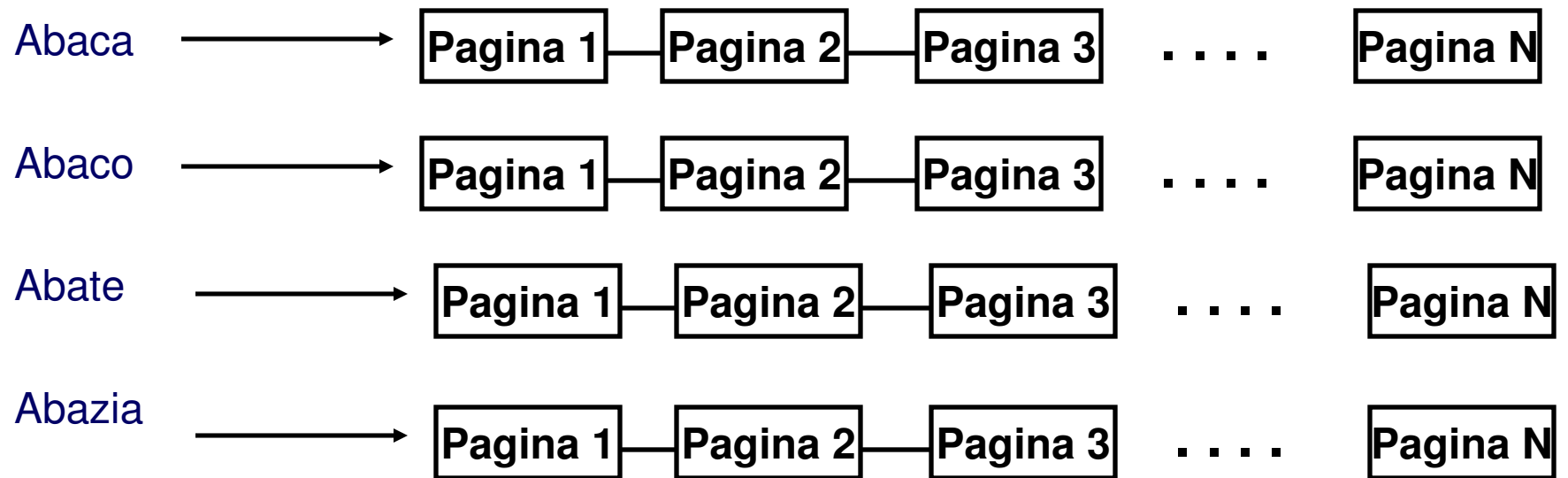
La Memorizzazione

- Il contenuto (testuale e altre informazioni) delle pagine analizzate viene memorizzato;
- La memorizzazione rende disponibile una copia dei **dati testuali** di ogni pagina archiviata per quando la risorsa originale è irraggiungibile: questa funzione è detta **Copia cache**;

Catalogazione

- Dalle pagine memorizzate vengono estratte le tutte le parole contenute (tranne preposizioni, articoli, ecc);
- Con queste parole si costruisce un **indice**;
- L'indice contiene in ordine alfabetico tutte le parole trovate nel web, e ad ogni parola è associato l'elenco delle pagine che la contengono;

L'indice



Risposta

- La risposta viene fornita cercando nell'indice le parole chiave fornite;
- Per ogni pagina trovata viene fornito:
 - *L'indirizzo della pagina;*
 - *Una parte del testo contenuto nella pagina (cache)*
 - *Un collegamento alla copia cache memorizzata*
- Le pagine trovate devono essere ordinate secondo dei criteri di rilevanza.

Classificazione delle Pagine

- Ogni motore di ricerca dei criteri per classificare le pagine;
- Alcuni esempi:
 - quante volte le parole chiave vengono ripetute nel documento;
 - quanti link riceve quel documento (PageRank di GOOGLE);
 - in quali punti della pagina si trovano le parole chiave;
 - quante volte un utente ha visitato quel sito dopo una ricerca;

Le dimensioni di un Motore di Ricerca

- Un motore di ricerca come GOOGLE è costituito da migliaia di servers (sistemi che siano circa 200 mila);
- Essi lavorano in maniera distribuita per:
 - Consentire a migliaia di utenti di accedere contemporaneamente;
 - Memorizzare (molte volte) l'enorme indice che contiene miliardi di parole diverse (divisi per lingua);
 - Memorizzare tutte le pagine analizzate (pagine cache)
 - Continuare ad analizzare in continuazione il Web (crawling)

Raffinare la Ricerca

- *La possibilità di raffinazione della ricerca varia da motore a motore, ma la maggior parte permette di utilizzare operatori logici;*

ES:

"Ganimede AND satellite NOT coppiere"

cerca informazioni su Ganimede inteso come pianeta e non come figura mitologica.

- *È possibile cercare contenuti presenti in un determinato sito,*

Es:

"Ganimede site:nasa.gov"

cerca le informazioni su Ganimede presenti sul sito della NASA.

Aggiungere il Proprio Sito

Google: Aggiungi la tua URL - Windows Internet Explorer

http://www.google.com/addurl/

Cerca con Google

Home Informazioni su Google

Aggiungi la tua URL

Ogni qual volta il nostro crawler scandaglia la rete, aggiungiamo nuovi siti al nostro indice. Ti invitiamo a segnalarci la tua URL. Non tutte le URL inoltrate saranno inserite nel nostro indice. Non è possibile prevedere o garantire se e quando esse saranno visibili.

Inserisci l'URL completa incluso il prefisso `http://`. Esempio: `http://www.google.com/`. È possibile inserire anche un commento o una parola chiave per descrivere il contenuto della tua pagina. Questi dati saranno utilizzati solo ad uso interno e non influiranno sul modo in cui la tua pagina verrà indicizzata ed usata da Google.

Nota: È sufficiente specificare il dominio di primo livello; non occorre inviare ogni singola pagina. Il nostro crawler, Googlebot, sarà in grado di trovare il resto.

Google aggiorna il proprio indice ogni qual volta è necessario; non occorre, pertanto, comunicare link aggiornati o non più validi. L'aggiornamento avverrà in modo automatico alla successiva scansione del crawler.

URL:

Commenti:

Facoltativo: Per aiutarci a distinguere URL inviati manualmente da URL inviati in modo automatico, inserisci la parola visualizzata nella casella qui in basso:

Internet 100%

Il Futuro dei Motori di Ricerca

Analisi semantica

A tutt'oggi le ricerche sono effettuate solo su testo, senza nessuna attribuzione di significato

In futuro i motori saranno capaci di distinguere il senso di una parola in base al contesto in cui è inserito.

Es: La Parola Madonna

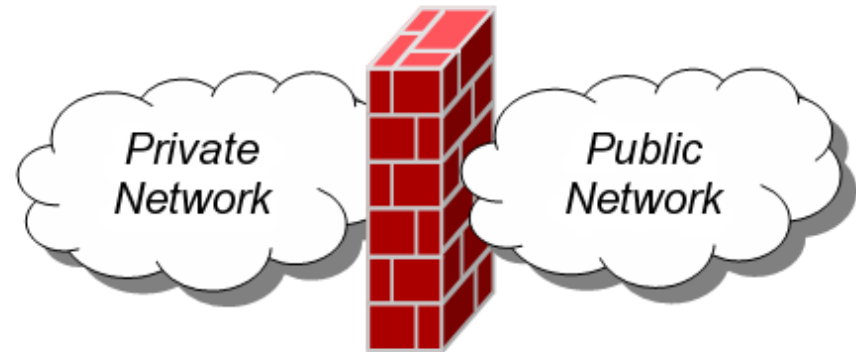
Si potrà distinguere tra i siti che parlano della madre di Gesù e quelli che parlano della cantante;

Contenuti Multimediali

In futuro sarà possibile anche effettuare ricerche non solo all'interno di documenti testuali, ma anche all'interno di immagini, video o audio

FireWall

- un **firewall** (muro tagliafuoco) è un componente passivo per la protezione di una rete LAN;
- Il firewall divide la rete in due sottoreti: una, detta esterna, comprende l'intera Internet mentre l'altra interna, è la LAN da proteggere
- In rete LAN il Firewall è un filtro dal quale transita il traffico di rete da e per l'esterno;
- Un firewall esamina tutti i pacchetti da e per l'esterno della rete;



Personal Firewall

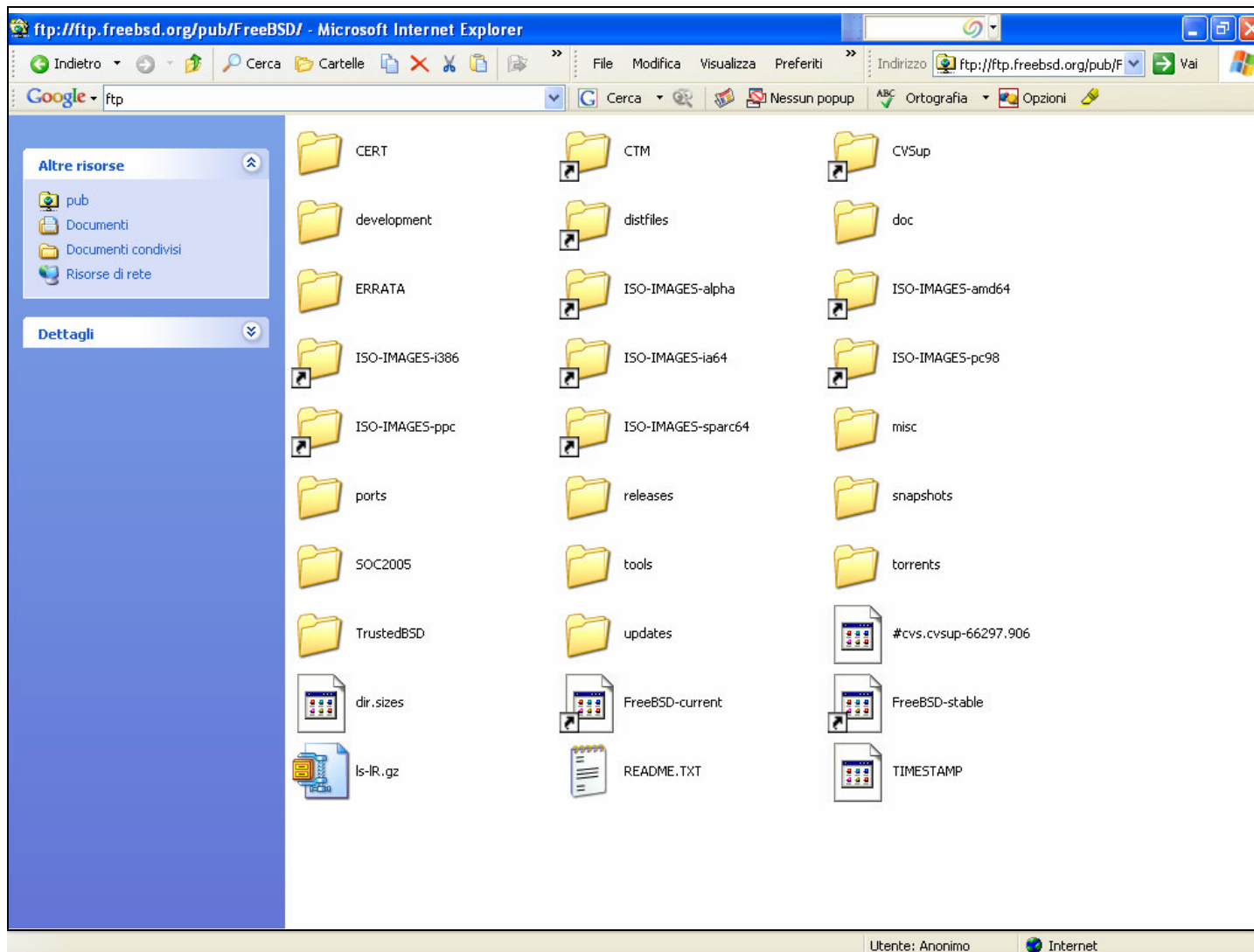
- Sui computer attuali sono disponibili dei personal Firewall;
- Un Personal Firewall è un programma che controlla tutto il traffico di rete che avviene su un singolo computer;
- Esso controlla tutti i programmi del computer che tentano di stabilire una connessione alla rete;
- Controlla anche i programmi che dall'esterno tentano di accedere alle risorse del computer;

Il protocollo FTP (1/3)

- Il File Transfer Protocol (FTP) è un particolare protocollo Internet che consente di accedere (scaricare) ai file presenti sulle macchine presenti in Internet.
- Il protocollo FTP presenta due modalità:
 - Senza autenticazione (Anonymous);
 - Con Autenticazione (Nome-utente e password);

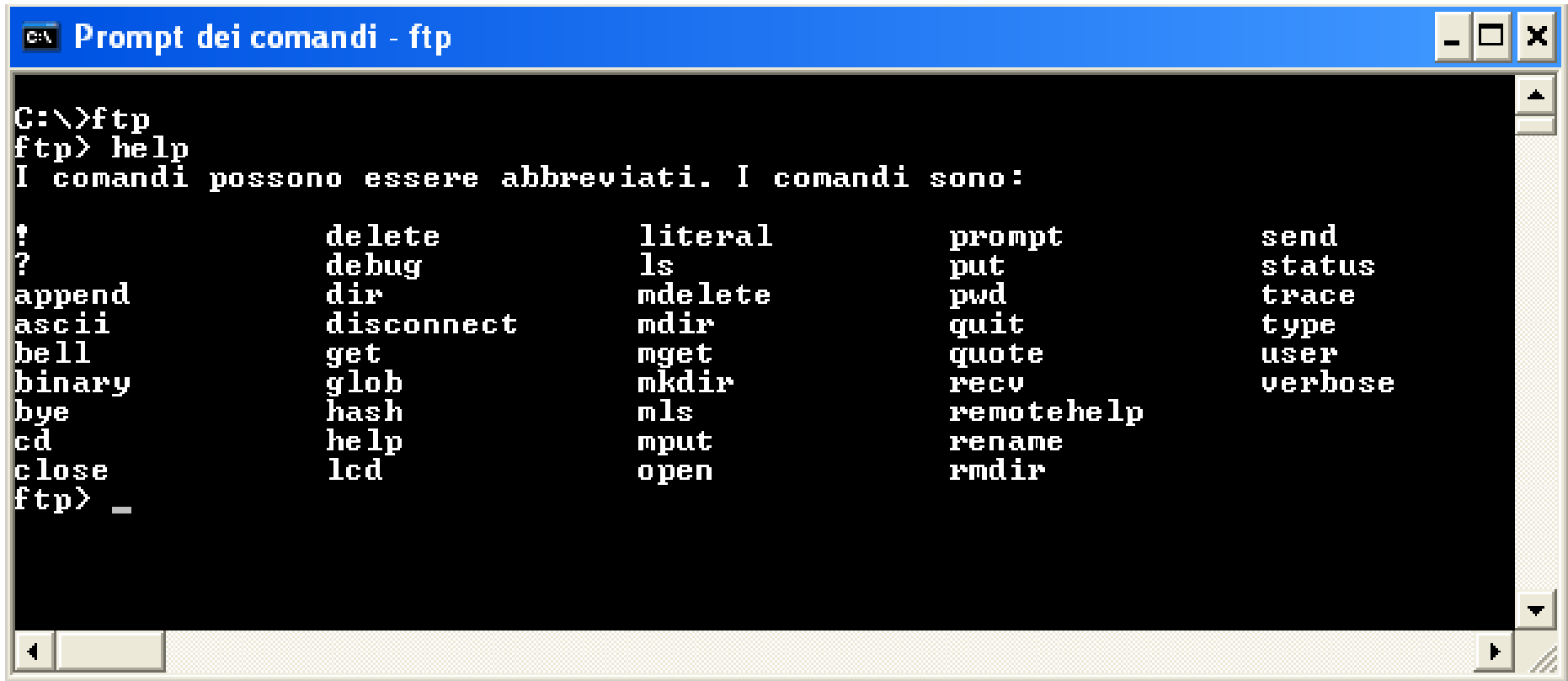
Il protocollo FTP (2/3)

L'accesso via FTP può essere fatto sia con il browser:



Il protocollo FTP (3/3)

L' FTP può essere fatto anche con uno specifico programma a linea di comando:



```
C:\>ftp
ftp> help
I comandi possono essere abbreviati. I comandi sono:
```

?	delete	literal	prompt	send
?	debug	ls	put	status
append	dir	mdelete	pwd	trace
ascii	disconnect	mmdir	quit	type
bell	get	mget	quote	user
binary	glob	mkdir	recv	verbose
bye	hash	mls	remotehelp	
cd	help	mput	rename	
close	lcd	open	rmdir	

```
ftp> _
```


La Posta elettronica (1/2)

- Il primo sistema di posta elettronica consisteva in semplici protocolli di trasferimento file;
- Per convenzione la prima linea contenesse l'indirizzo del destinatario;
- Non era possibile inviare in maniera semplice immagini o audio

La Posta elettronica (2/2)

- I moderni sistemi di posta elettronica sono costituiti da due sottosistemi:
 - Agenti utente
 - Agenti di trasferimento messaggi;
- I primi si occupano della preparazione del messaggio, mentre i secondi del suo trasferimento;

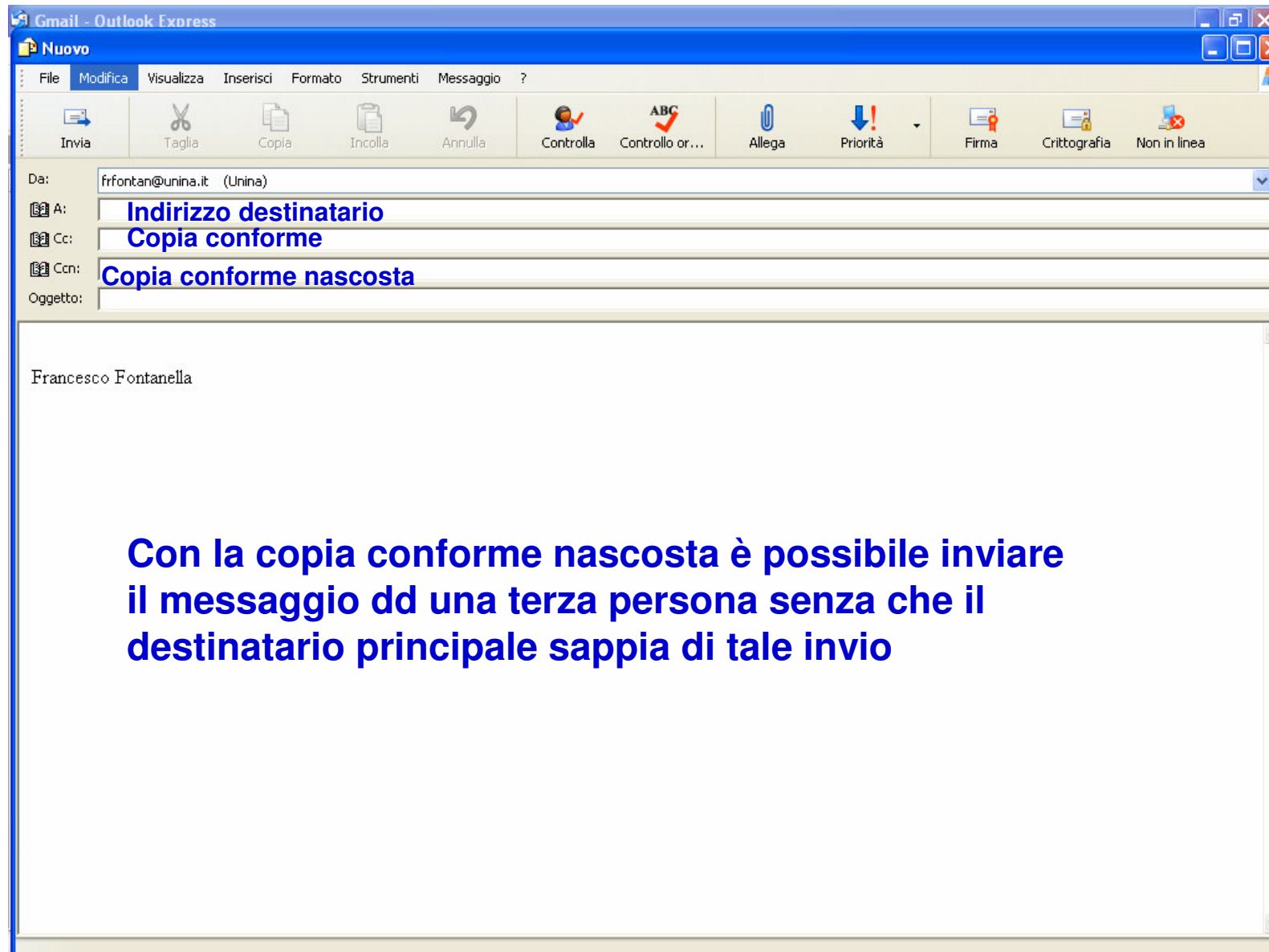
Il formato dei messaggi

- Oggi i messaggi di posta elettronica sono preparati secondo il formato:

MIME (Multipurpose Internet Mail Extensions)

- Questo protocollo prevede l'intestazione e il corpo del messaggio;
- Il corpo del messaggio può essere di vari tipi:
 - Testo;
 - Jpeg, Mpeg;
 - Binario in genere

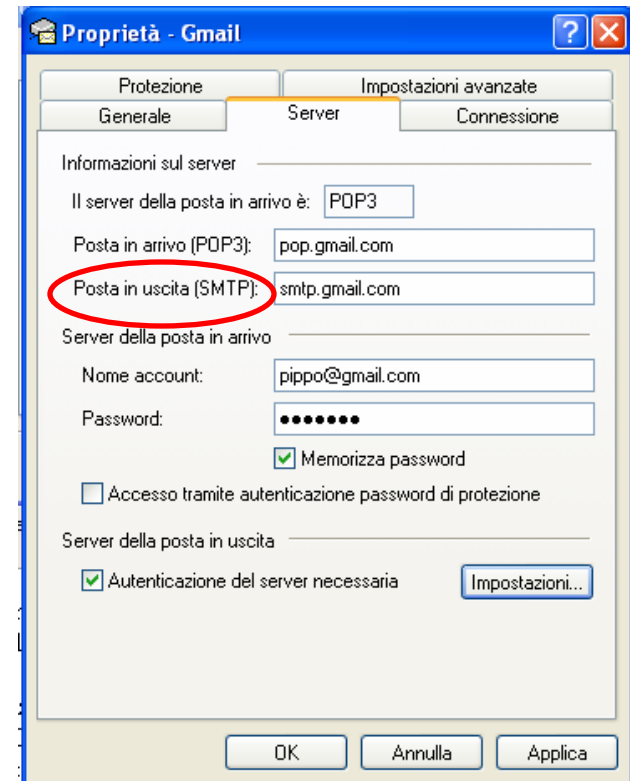
Invio di un Messaggio



Con la copia conforme nascosta è possibile inviare il messaggio da una terza persona senza che il destinatario principale sappia di tale invio

SMTP

- **Simple Mail Transfer Protocol (SMTP)** è un protocollo che consente il trasferimento di messaggio tra le macchine in rete, esso usa la porta 25 del protocollo TCP;
- Sui server di posta elettronica, su questa porta è in ascolto un demone (processo server) che aspetta l'invio di messaggi;



POP

- Post Office Protocol (POP3) è un protocollo che consente di recuperare la posta presente su una casella di posta elettronica remota e memorizzarla sulla macchina locale dell'utente;
- Anche in questo caso sul server c'è un demone in ascolto (porta 110) che attende richieste da parte di clients per inviargli i messaggi memorizzati;

