

Introduzione a Internet e World Wide Web

Sommario

- **Breve storia di Internet**
- **Commutazione di pacchetto e TCP/IP**
- **Il Web**
 - HTTP**
 - HTML**
 - CGI**
 - ...**
- **Connessione tra basi di dati e Web**

Internetworking

(collegamento fra reti e sistemi eterogenei)

- **Nasce da un progetto militare (DARPA)
→ ARPAnet (1975-80)**
- **Con l'aggregazione di laboratori di ricerca e università USA diventa Internet**
- **Sviluppo:**
 - **reti scientifiche nazionali e continentali**
 - **aggregazione reti private / commerciali**
- **Salto di qualità (anni '90):
Tecnologia World Wide Web**

Internet: una definizione

Sistema globale di informazione che:

- è collegato da un unico spazio di indirizzi basato sul protocollo IP**
- consente di comunicare mediante protocollo TCP/IP o sue estensioni**
- rende accessibili, pubblicamente o privatamente, servizi basati sulla infrastruttura descritta**

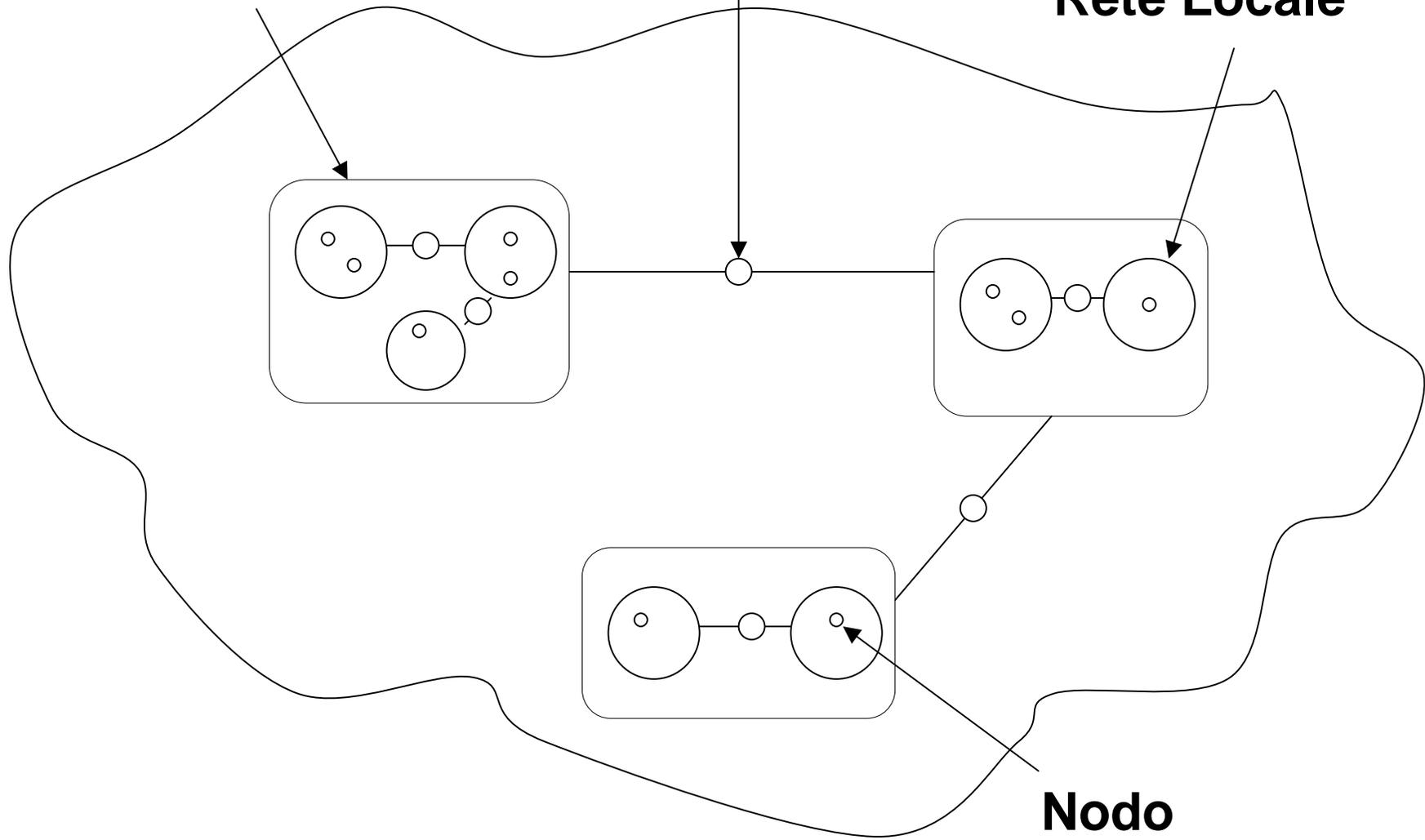
***Federal Networking Council,
24/10/1995***

Internet come federazione di reti

Rete Componente

Router

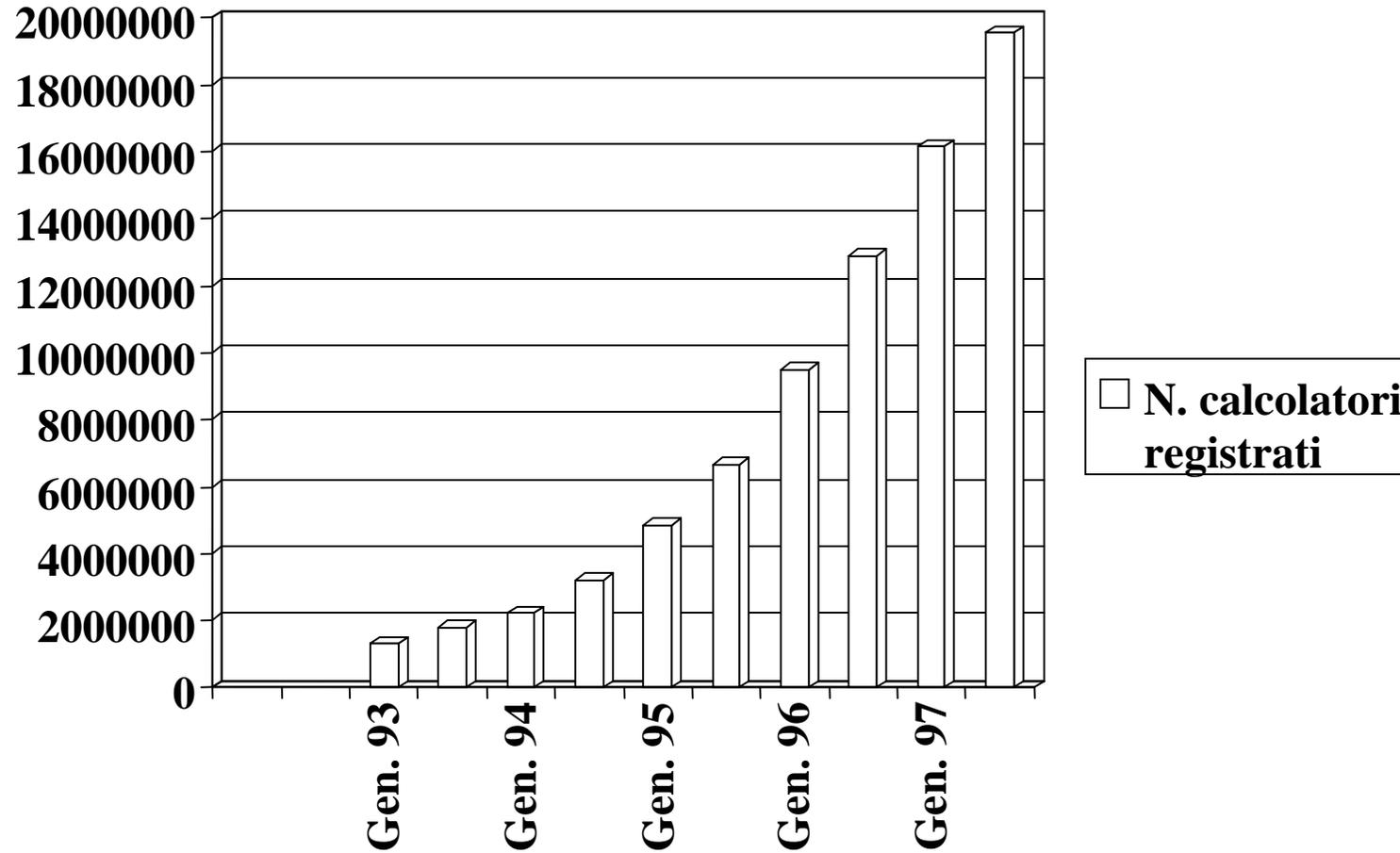
Rete Locale



Internet come federazione di reti

- **Ogni nodo ha un indirizzo numerico (Indirizzo IP)
es. 137.204.57.150**
- **Ad ogni nodo è associato un nome simbolico utilizzato nelle applicazioni per riferirsi ad esso
es. www-db.deis.unibo.it**
- **La conversione da nome simbolico ad indirizzo IP è gestita dai Domain Name Server (deis.unibo.it è un nome di dominio)**

Lo sviluppo di Internet



World Wide Web: una definizione

Collezione distribuita di documenti ipertestuali e multimediali disponibili su Internet.

Alla base di WWW c'è:

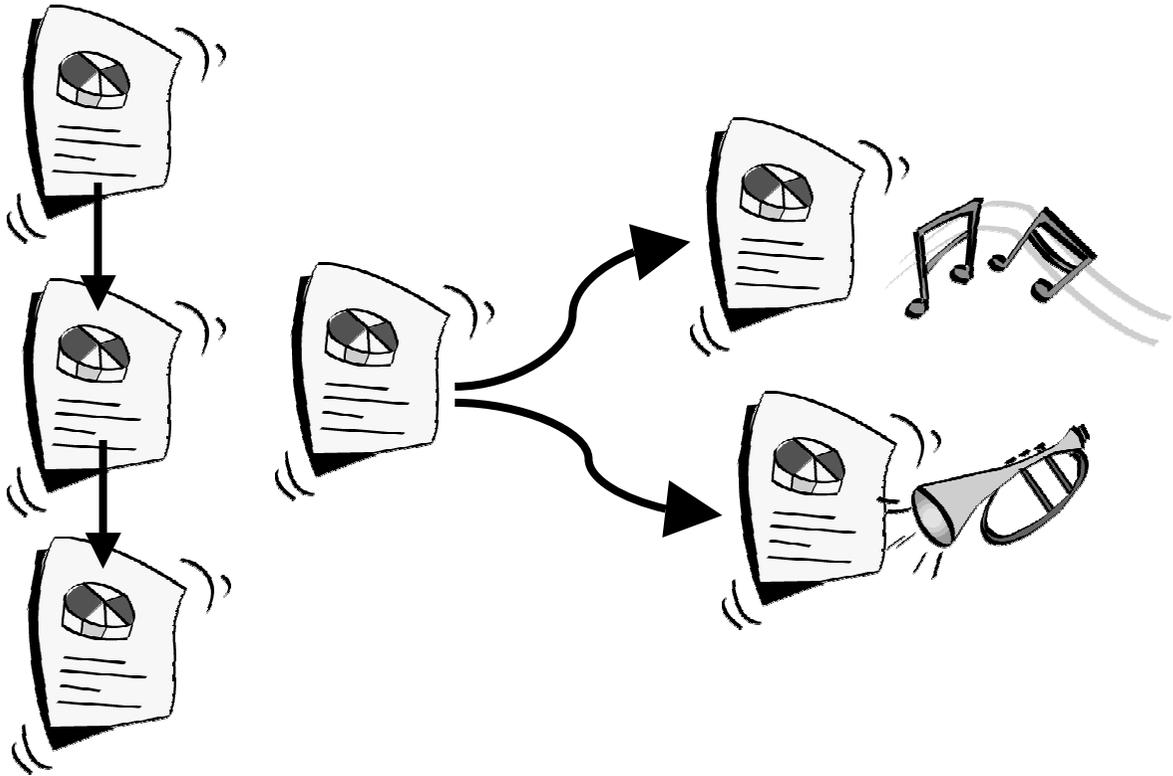
- Uno standard per identificare risorse di rete e poterle specificare all'interno di documenti ipertestuali (URL - Uniform Resource Locator)**
- Uno o più linguaggi per la realizzazione di documenti ipertestuali (HTML - HyperText Markup Language, XML...)**
- Un protocollo di comunicazione client-server per lo scambio di documenti ipertestuali (HTTP - HyperText Transfer Protocol)**

Iper testo (multimediale)

Documento con una struttura non sequenziale, costituito da varie porzioni fra loro collegate, al fine di favorirne la consultazione a seconda delle esigenze.

Può contenere:

- **testo**
- **immagini**
- **video**
- **suoni**
- **ecc.**



Protocolli e linguaggi di Internet

TCP / IP, HTTP, HTML

Il protocollo TCP/IP

- **Transfer Control Protocol / Internet Protocol**
- **Insieme di protocolli (suite) strutturati in maniera gerarchica adibiti alla comunicazione (in Internet)**
- **Ogni livello della gerarchia (detta PILA) usa i servizi forniti dal livello inferiore e offre servizi al livello superiore**
- **Fra livelli contigui è definita una interfaccia di comunicazione**

La pila TCP/IP

application layer	offre servizi utilizzabili dalle applicazioni o dall'utente (es. telnet, ftp, smtp, http)
transport layer	fornisce meccanismi di comunicazione tra applicazioni (TCP, UDP)
internet layer	fornisce il meccanismo di trasmissione dati attraverso la rete Internet (IP)
network interface layer	nasconde i dettagli di accesso alla rete fisica ai livelli superiori

TCP e IP

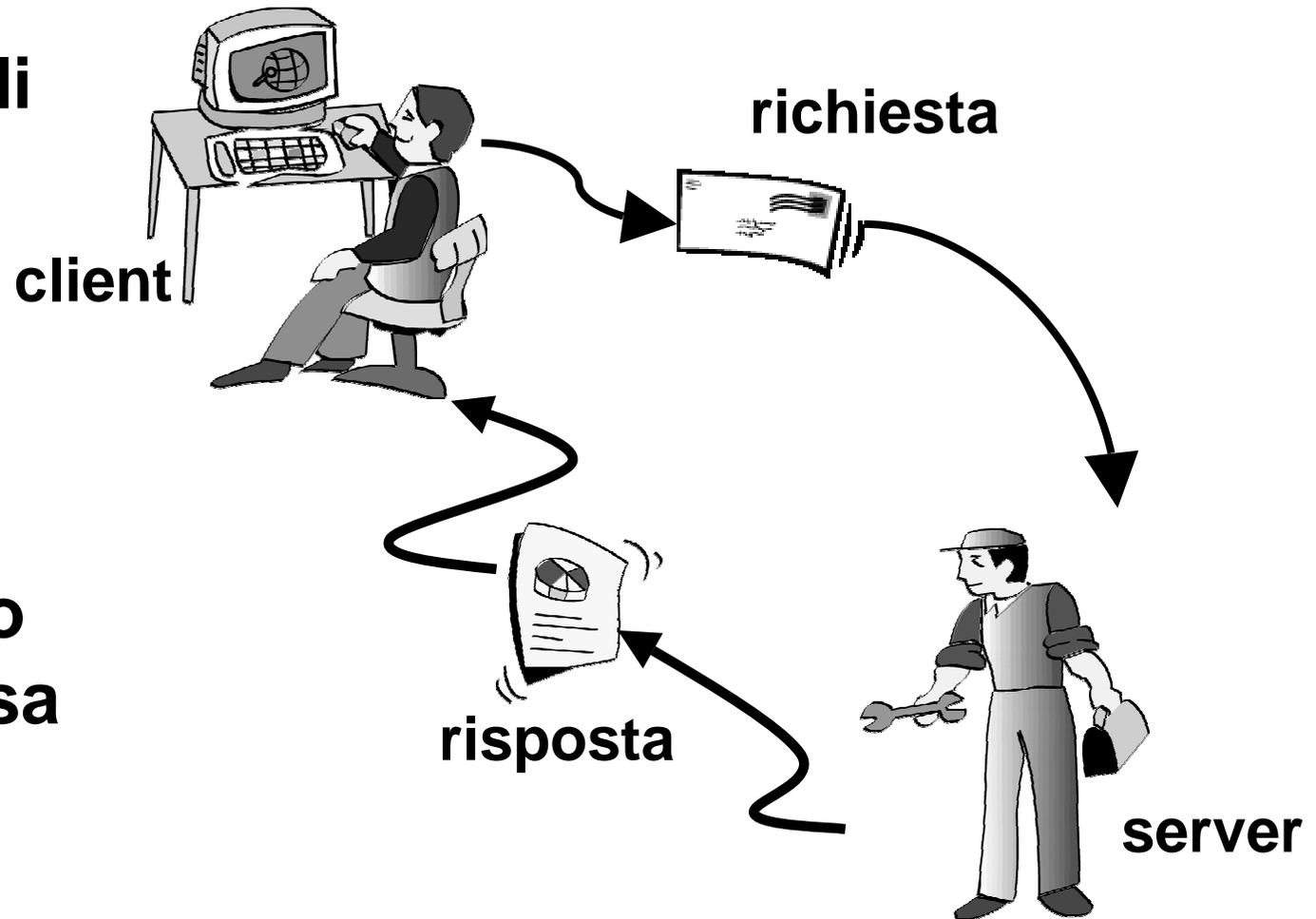
- **Internet Protocol**
 - **si occupa della spedizione in rete di unità di informazione (datagram)**
 - **definisce il meccanismo di indirizzamento dei nodi di Internet (IP address)**
- **Transfer Control Protocol**
 - **converte il flusso dati proveniente dalla applicazione in pacchetti (segment) trasmissibili in rete**
 - **verifica integrità e correttezza dei dati trasmessi (recupero/ritrasmissione pacchetti persi, rimozione duplicati ...)**

Il protocollo HTTP

- **HyperText Transfer Protocol**
- **Protocollo a livello di applicazione per lo scambio di ipertesti multimediali**
- **Prescrive il formato di**
 - **nomi delle risorse (URI, URL)**
 - **domande**
 - **risposte**
- **Versioni: HTTP/0.9, 1.0, 1.1**
- **Riferimento: Tim Berners Lee, Request for Comment 1945, HTTP/1.0**

Gli attori di HTTP

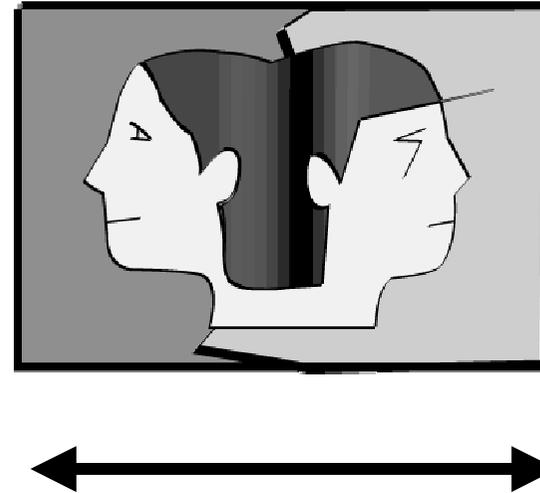
- **client**
(browser):
genera
richieste di
risorse



- **server:**
depositario
della risorsa
richiesta

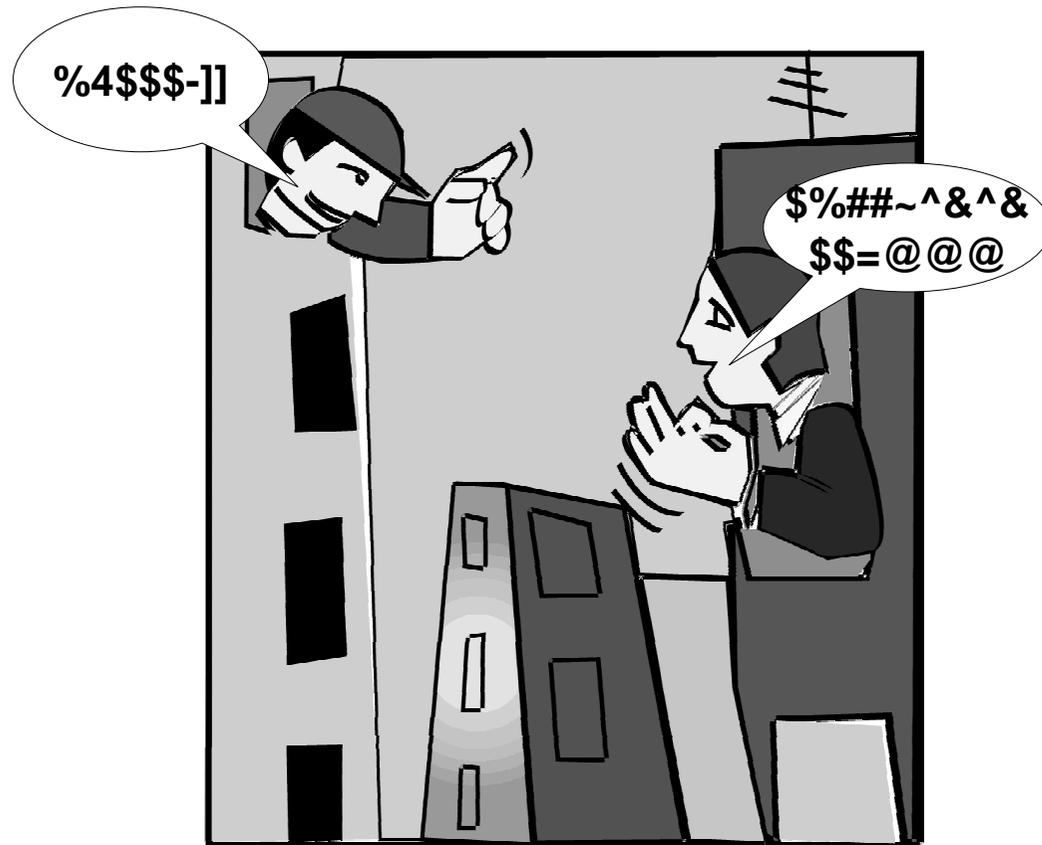
Gli attori di HTTP

- **proxy:**
nodo
intermedio
che può
fungere sia
da client che
da server



Gli attori di HTTP

- gateway:
nodo
intermedio
che può
agire solo
da server,
per tradurre
richieste
in altri
protocolli
o a fini di
protezione



Uniform Resource Locator (URL)

- E' una stringa strutturata, ad esempio:
`http://www-db.deis.unibo.it/~fgrandi/index.html`
- Protocollo: http, ma anche ftp e gopher
- Indirizzo della macchina:
 - simbolico: `www-db.deis.unibo.it`
 - numerico (IP): `137.204.57.150`
 - può includere il numero di porta (es. `:8080`)
- Path: sequenza di direttori
- Nome risorsa: identificativo di un file
 - se la risorsa e' un file html, puo' includere un indirizzo interno, es. `index.html#curriculum`

Formato delle richieste

- **full-request :- request-line**
 - *(general-header | request-header | entity-header)**
 - CRLF [entity-body]**
- **request-line :-**
 - method SP URL SP version CRLF**
- **method :- GET | POST | HEAD | altri..**
- **Esempio:**
 - GET /pub/papers/pap101.html HTTP/1.0**

Formato delle risposte

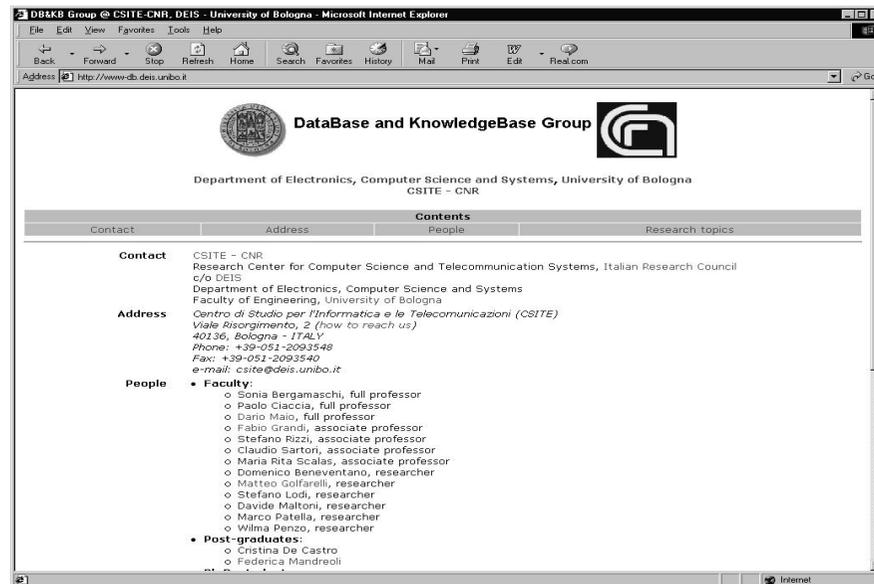
- **full-response :- status-line**
 - *(general-header | request-header | entity-header)**
 - CRLF [entity-body]**
- **status-line :-**
 - version SP status SP message CRLF**
- **Codici di stato:**
 - 1XX (informativo), 2XX (successo),**
 - 3XX (redirezione), 4XX(errore client),**
 - 5XX (errore server)**

Formato dei documenti

- I documenti risiedono sul server come file e vengono inviati al client usando il campo **entity-body** della risposta
- **Es: Iper testo = file ASCII scritto secondo la sintassi HyperText Markup Language (HTML)**
- **Estensioni tipiche del file: *.html *.htm**
- **Il documento può contenere inserti multimediali (immagini, audio, video)**
 - **il programma client carica e visualizza il testo**
 - **poi richiede al server i file corrispondenti agli inserti multimediali**

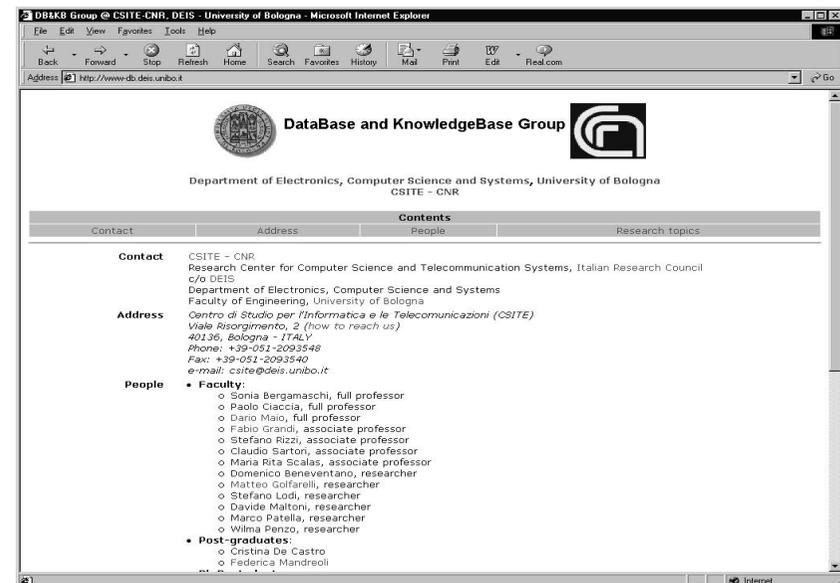
Il Browser (client HTTP)

- Applicazione in grado di:
 - accedere alla rete secondo il protocollo HTTP
 - richiedere risorse identificate da un URL a un server
 - interpretare e rendere a video la risposta del server
 - gestire la navigazione (click su link)



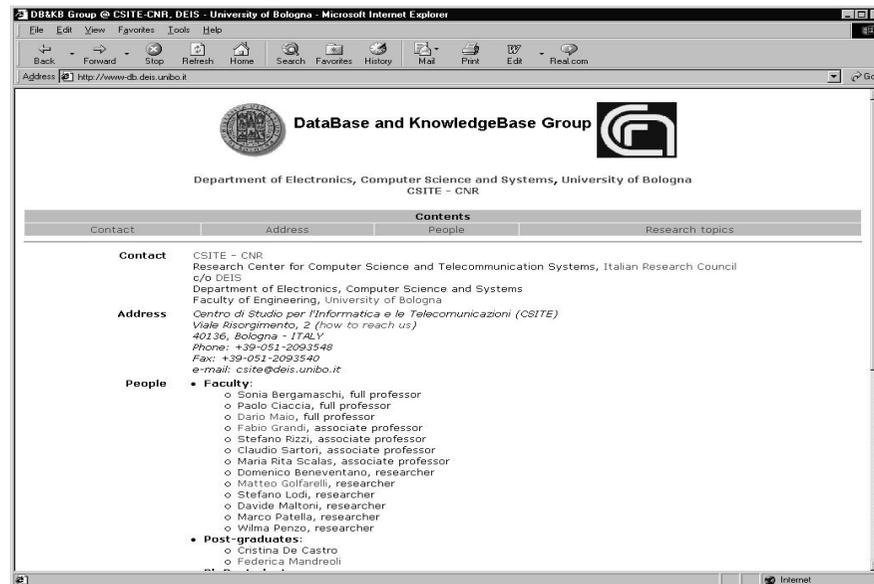
Il browser Web

- I diversi prodotti differenziano per:
 - La versione di HTML trattata (nuove versioni supportano anche XML)
 - La capacità' di trattare estensioni non standard di HTML (ad es. JavaScript, VBscript)
 - La capacità di eseguire programmi (es. Java)
 - Netscape ed Explorer sono i più diffusi



Il Browser Web

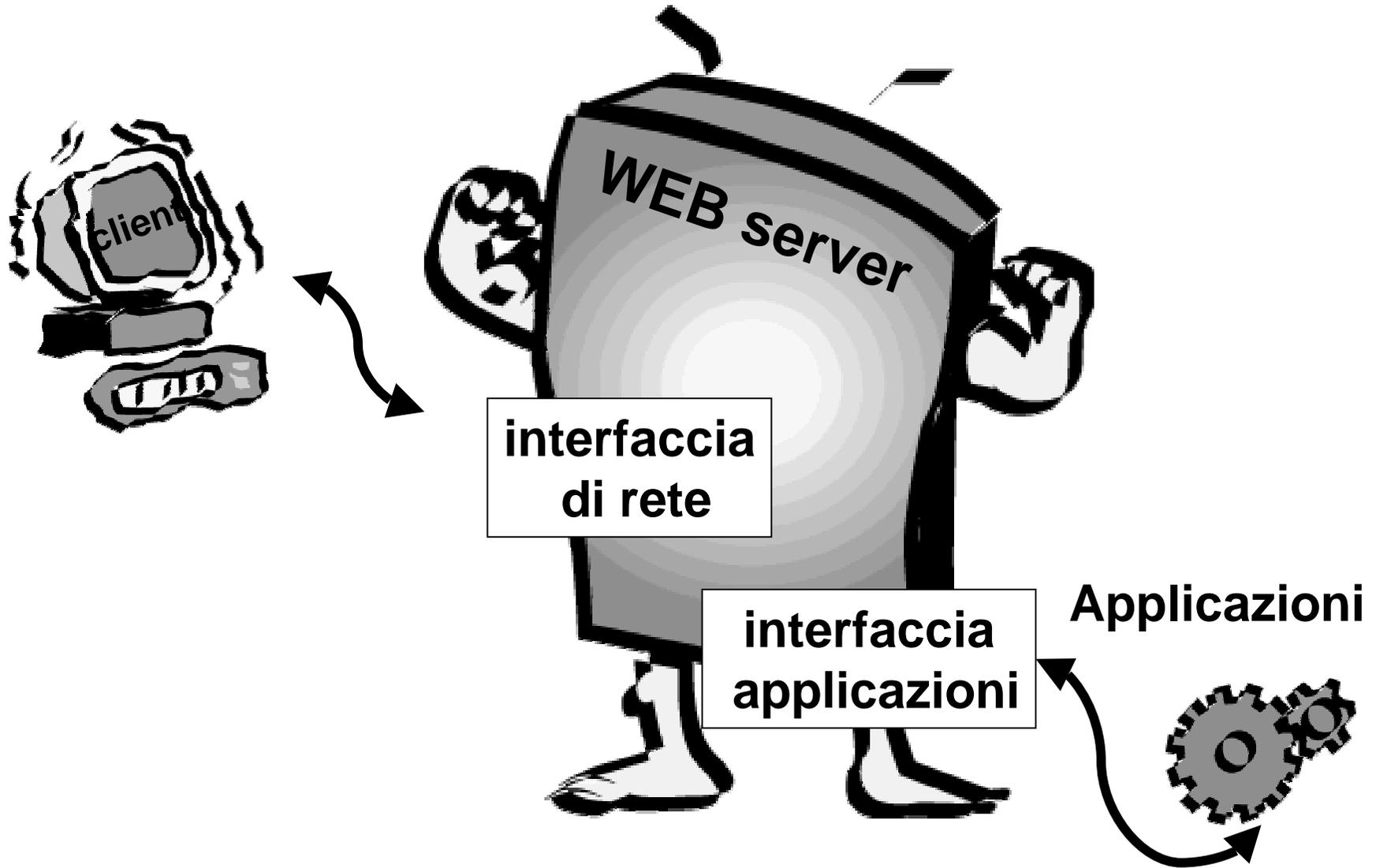
- Le capacità possono essere ampliate con l'aggiunta di applicazioni esterne (plug-in):
 - per trattare audio, video, animazioni
 - per visualizzare interfacce 3D (ed es. VRML)
 - per eseguire applicazioni gestionali (es. PowerBuilder)



Il server HTTP

- **Funzioni base:**
 - **accesso alla rete secondo il protocollo HTTP**
 - **invio di risorse identificate da un URL a un client**
 - **controllo degli accessi**
 - **lancio di programmi in risposta a richieste**
 - **registrazione degli accessi (logging)**
- **Funzioni avanzate:**
 - **monitoraggio e amministrazione**
 - **connessione a basi di dati**
 - **esecuzione efficiente di applicazioni esterne**

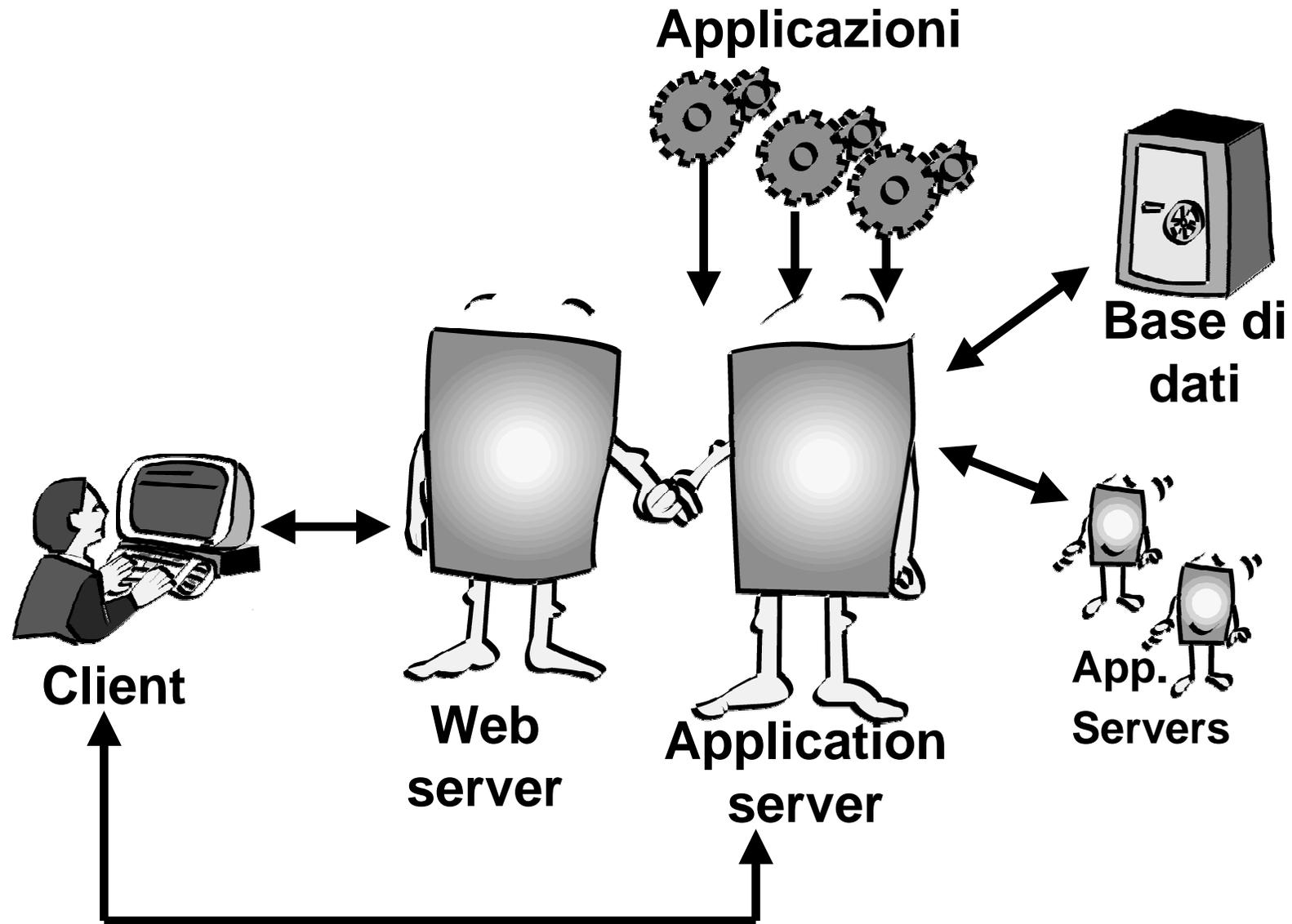
Server HTTP: architettura base



Architettura base: sessione

- **HTTP/1.0 prevede una diversa connessione tra client e server per ogni ciclo domanda/risposta**
- **Il server HTTP/1.0 chiude la connessione dopo l'invio della risposta**
- **Una seconda richiesta dello stesso client richiede l'apertura di una nuova connessione**
- **La nuova connessione non ricorda nulla della precedente (stateless)**

Server HTTP: architettura avanzata

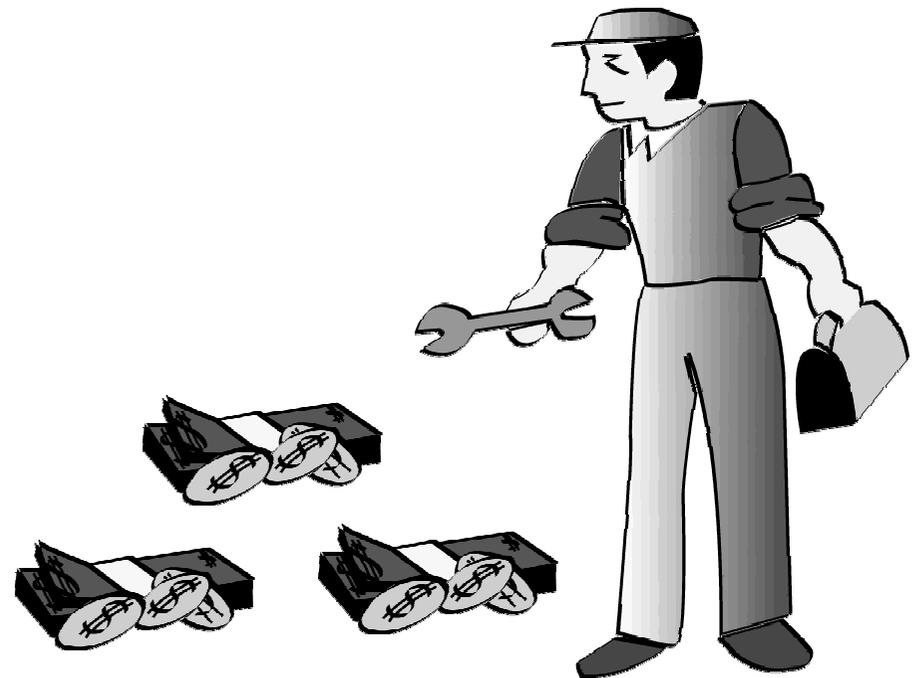


Architettura avanzata: vantaggi

- **L'application server può**
 - **registrare i clienti attivi**
 - **memorizzare informazioni di contesto tra connessioni diverse**
(globalmente, per cliente, per sessione)
 - **ottimizzare le connessioni a basi di dati**
- **per aumentare le prestazioni, si può usare**
 - **multi-threading, multi-processing, distribuzione su diverse macchine**
- **oltre ad HTTP, si possono gestire anche altri protocolli (Corba IIOP, COM/DCOM)**

Sicurezza in HTTP/1.0

Le risorse del server sono organizzate in distinti domini (realms)

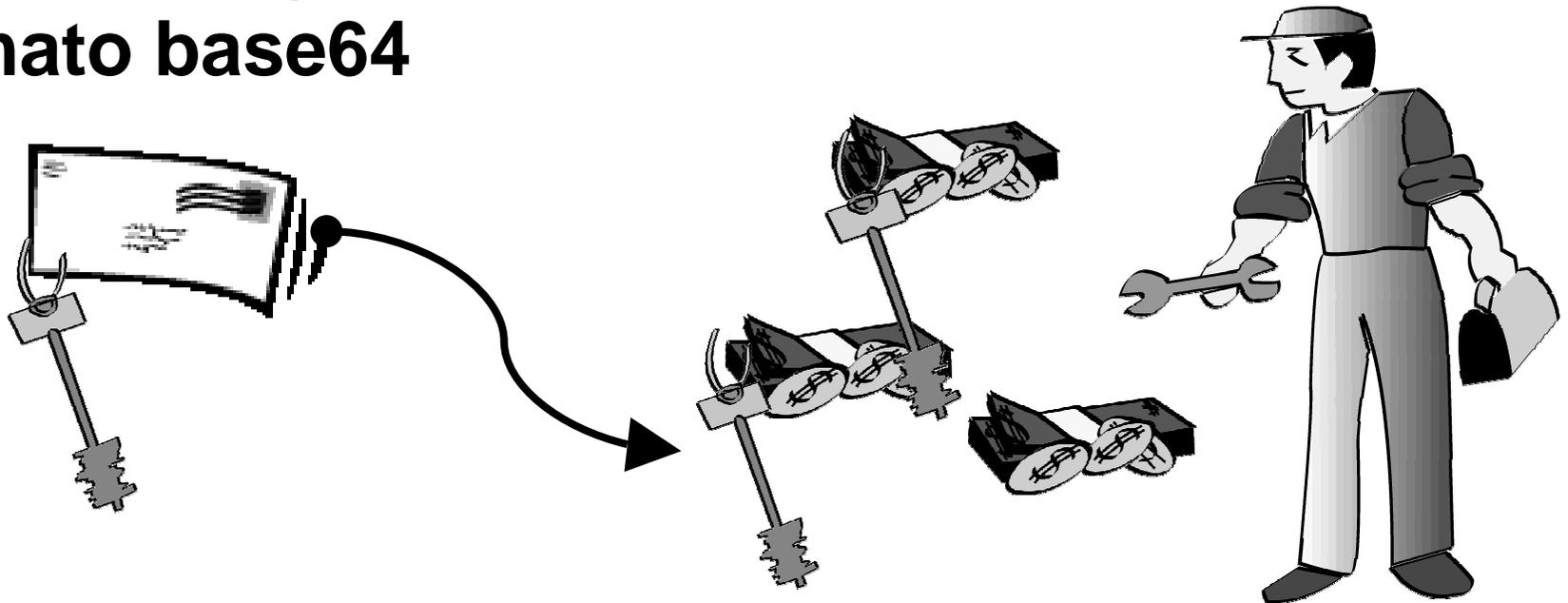


Sicurezza in HTTP/1.0

Richieste di risorse protette devono contenere un campo (authorization)

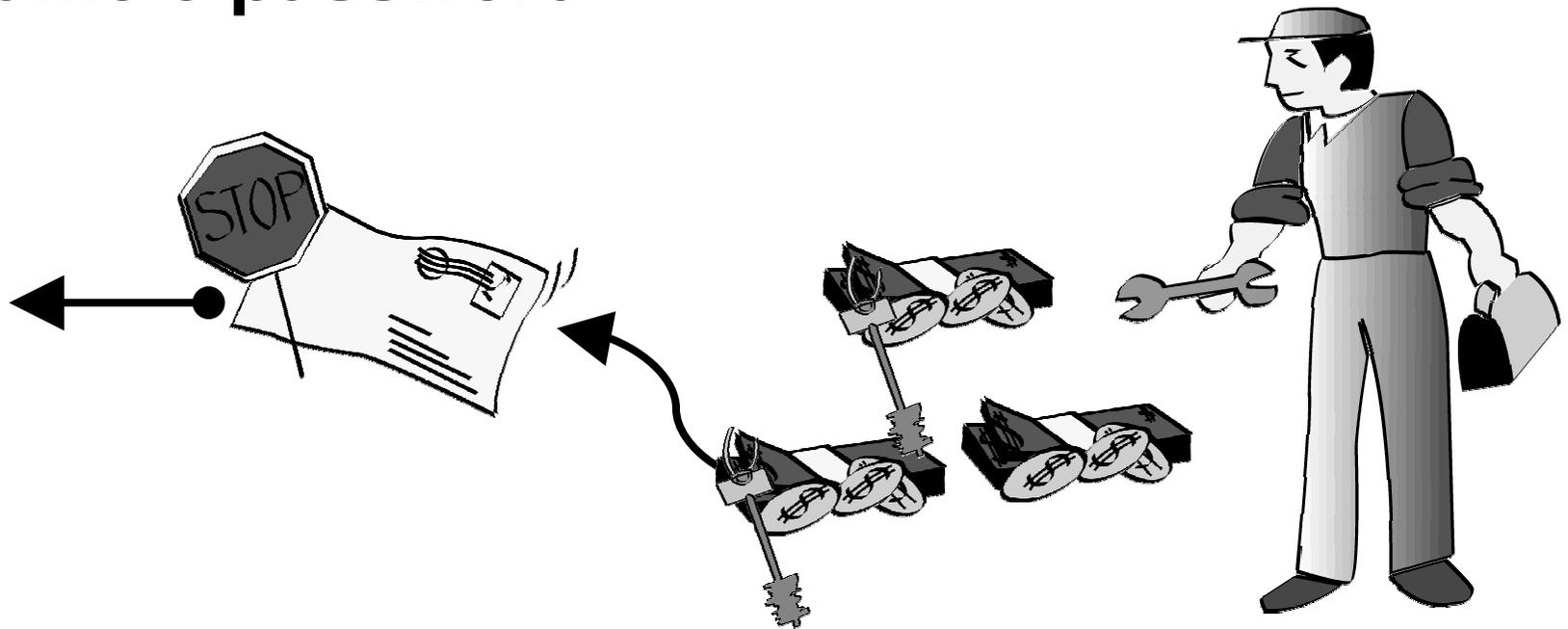
con le credenziali (username + password)

NB: le credenziali vengono trasmesse sotto forma di stringa codificata secondo il formato base64



Sicurezza in HTTP/1.0

Se la richiesta non contiene le giuste credenziali, il server risponde con il codice 401 (unauthorized) + campo (authenticate) che provoca la richiesta di username e password



HyperText Markup Language

- **Linguaggio di descrizione di testi secondo lo schema SGML (Standard General Markup Language)**
- **I documenti HTML possono essere visualizzati per mezzo di browser**
- **Un documento comprende un contenuto e una marcatura che definisce la forma (formattazione) del contenuto**

Concetti generali di HTML

- **La marcatura prevede l'uso di etichette, dette TAGS**
- **I tag viaggiano (quasi) sempre in coppia**
 - **<tag> testo marcato </tag>**
- **Il significato di un tag può essere modificato tramite attributi**
 - **<tag attributo=valore> testo </tag>**

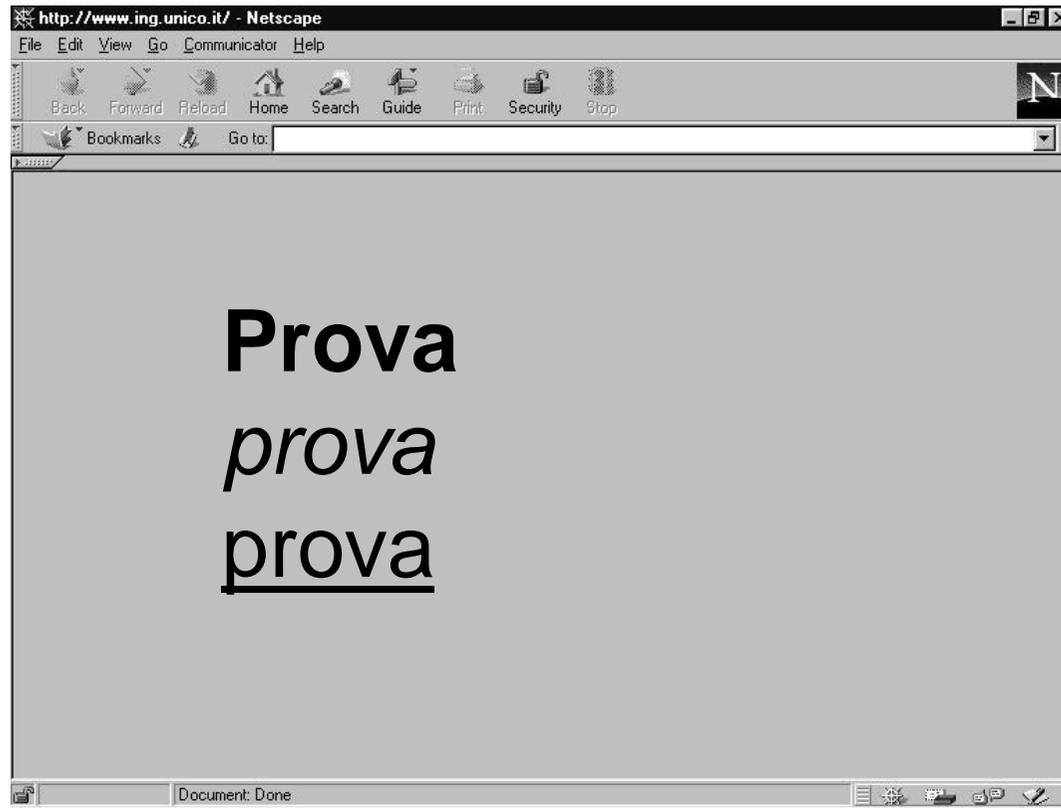
Struttura del documento

- **File ASCII, struttura generale:**
<html> intestazione + corpo </html>
- **Intestazione: <head> .. </head>**
contiene informazioni sul documento:

es. titolo <title> .. </title>
- **Corpo: <body> .. </body>**
 - **contiene il testo del documento**
e i tag per la resa visiva

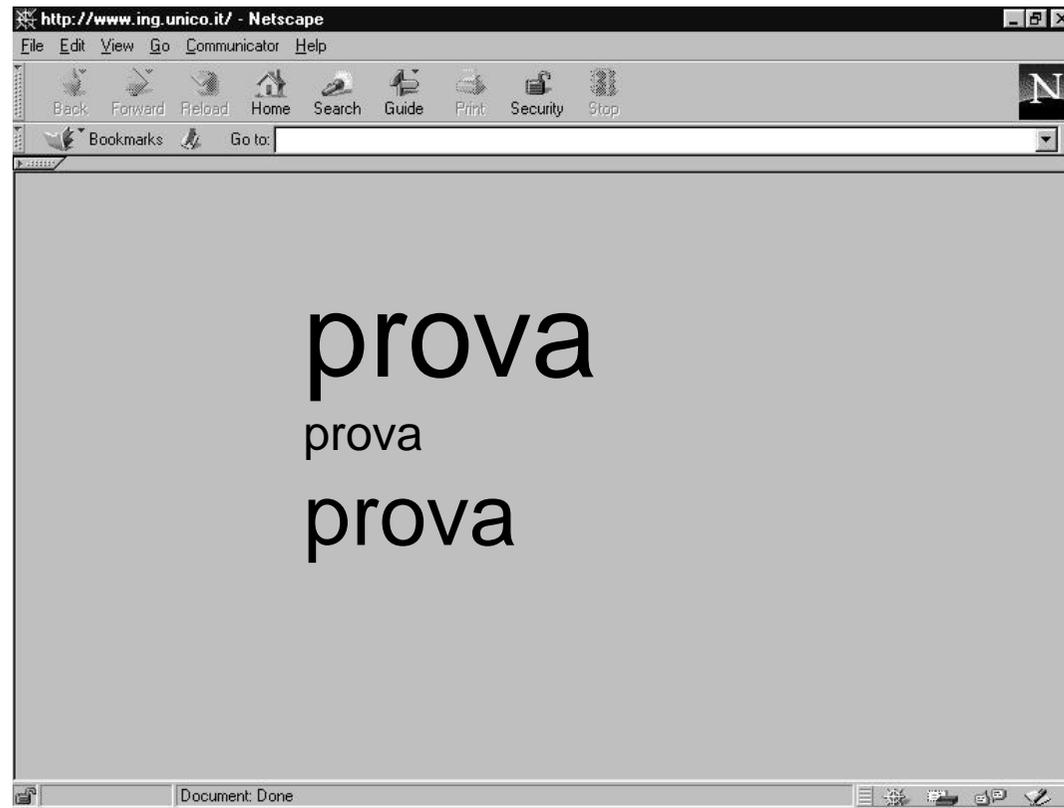
Formattazione del testo

- **grassetto** `` prova ``
- **corsivo** `<i>` prova `</i>`
- **sottolineato** `<u>` prova `</u>`



Formattazione del testo

- **Dimensioni:** ` prova`
` prova`
- **Colore:** ` prova`



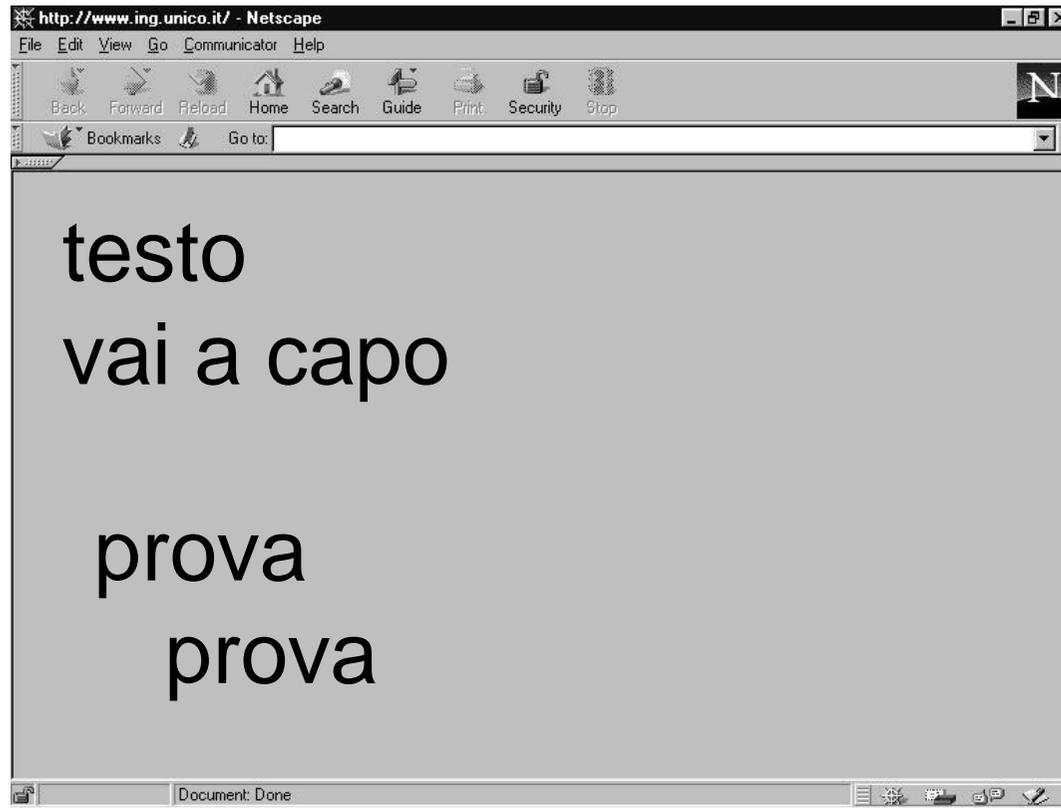
Titoli

- **titolo 1** `<h1>Titolo livello 1</h1>`
- **titolo 2** `<h2>Titolo livello 2 </h2>`
- **titolo 3** `<h3>Titolo livello 3 </h3>`



Paragrafi, allineamenti

- paragrafi: `<p>testo</p>` vai a capo `
`
- testo formattato: `<pre>prova
prova</pre>`



Liste

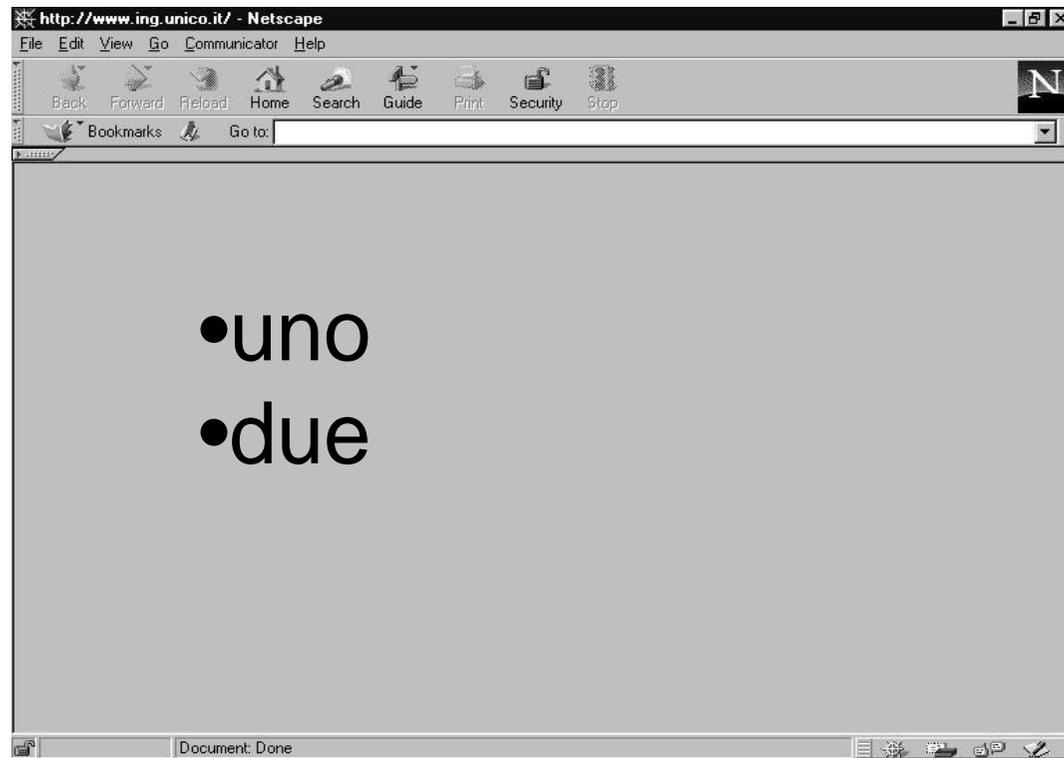
``

` uno `

` due `

``

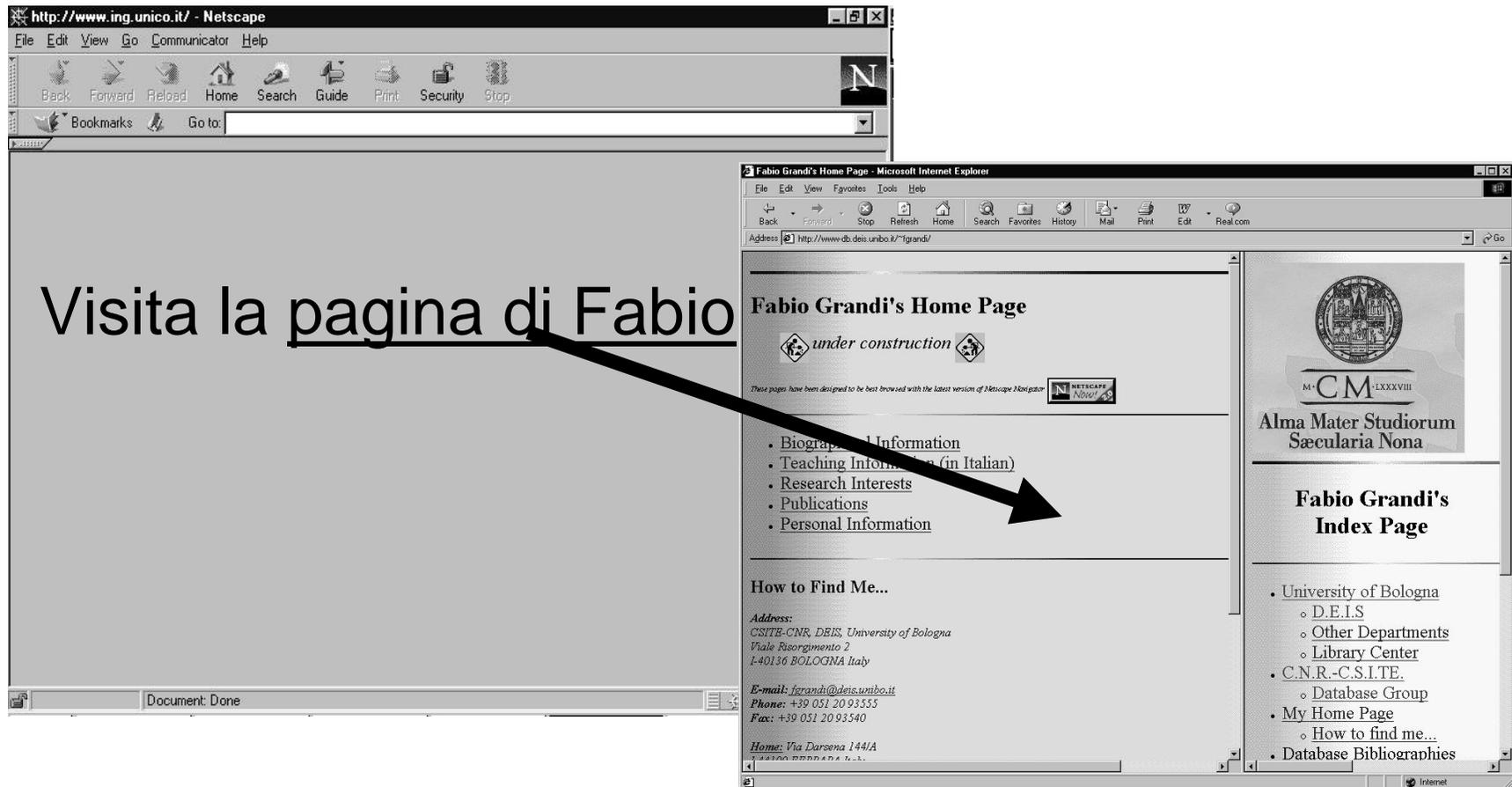
**NB: per liste numerate
si usa ``**



Collegamenti ipertestuali

- **Esempio di collegamento : Visita la**

pagina di Fabio



Collegamenti ipertestuali

- **Uso della posta:**

Scrivi a``

Fabio Grandi``



Lancia il programma di posta elettronica preimpostato (MAPI server)

Immagini

```
<p align=center >Ecco la mia foto: </p>  
<img src = "Fabio.gif" width=200 height=400  
alt="Foto di Fabio">
```



Immagine+collegamento

```
<p align=center >Clicca sulla foto per scrivermi: </p>  
<a href = mailto://fgrandi@deis.unibo.it >  
<img src = "Fabio.gif" width=200 height=400  
alt="Foto di Fabio"> </a>
```

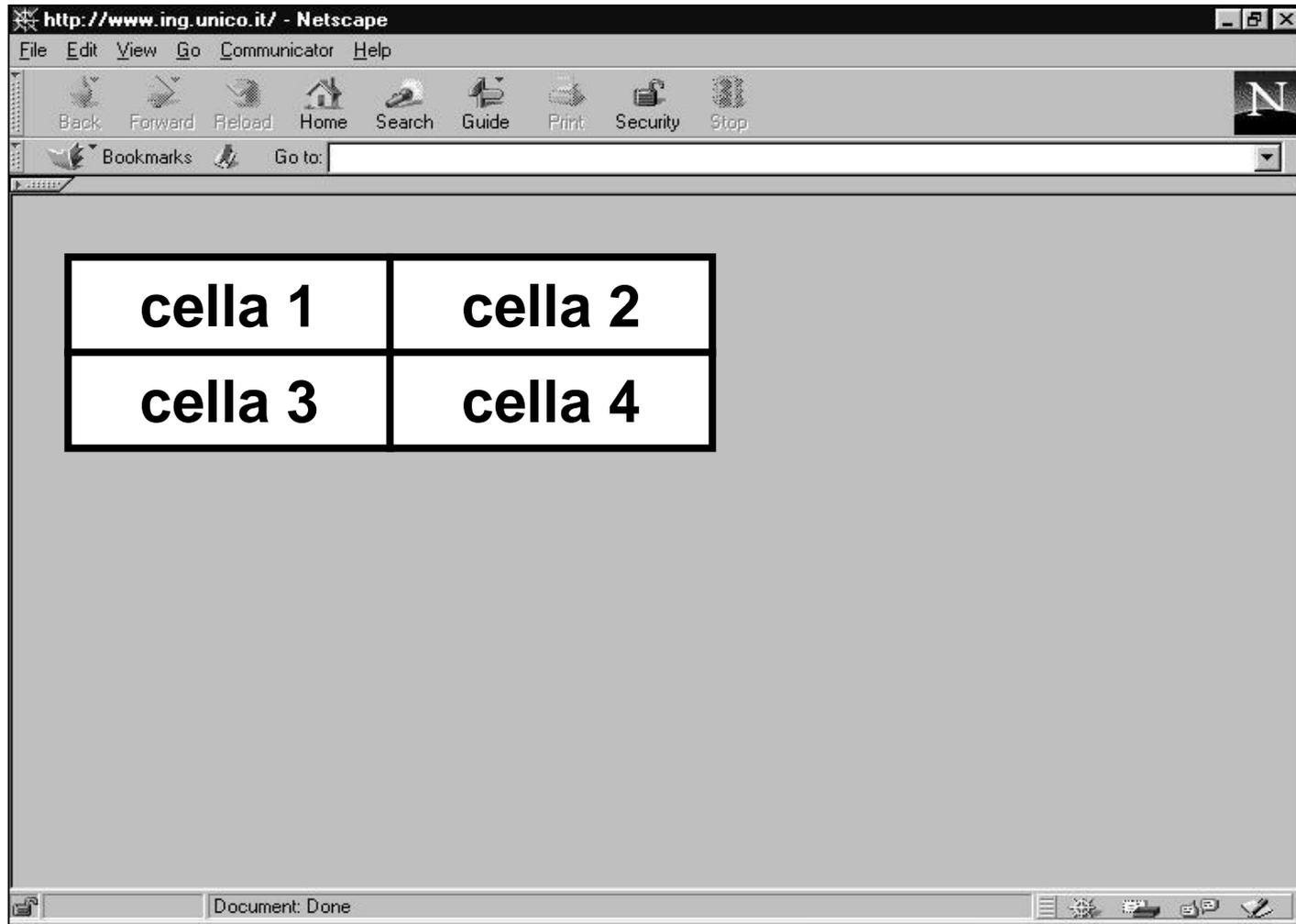


Attiva programma
di posta elettronica

Tabelle

```
<table border="1" width="50%"
  bgcolor="#C0C0C0">
  <tr>
    <td width="50%" bgcolor="#00FF00">
      cella 1</td>
    <td width="50%">
      cella 2</td>
  </tr>
  <tr "bgcolor="#00FFFF">
    <td width="50%">cella 3</td>
    <td width="50%">cella 4</td>
  </tr>
</table>
```

Tabelle



Limiti di HTML

- **HTML consente di descrivere documenti ipertestuali statici**
- **Questo non consente di:**
 - **costruire un documento a partire da dati contenuti nel file system o in una base di dati (ad es. una bacheca di messaggi)**
 - **personalizzare un documento in funzione di chi lo richiede**
- **Esigenza: costruire un documento HTML dinamico “su richiesta”**