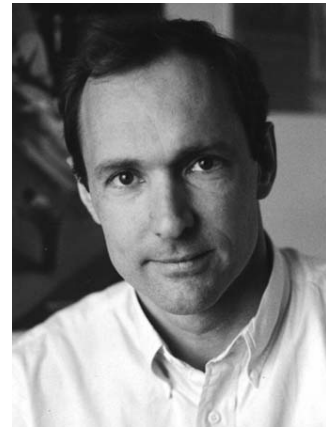


World Wide Web

Quando e perché nasce

Tim Berners-Lee (1989, CERN di Ginevra)

- “The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, ...”
- “A link is specified as an ASCII string from which the browser can deduce a suitable method of contacting an appropriate server. When a link is followed, the browser addresses the request for the node [document] to the server.”



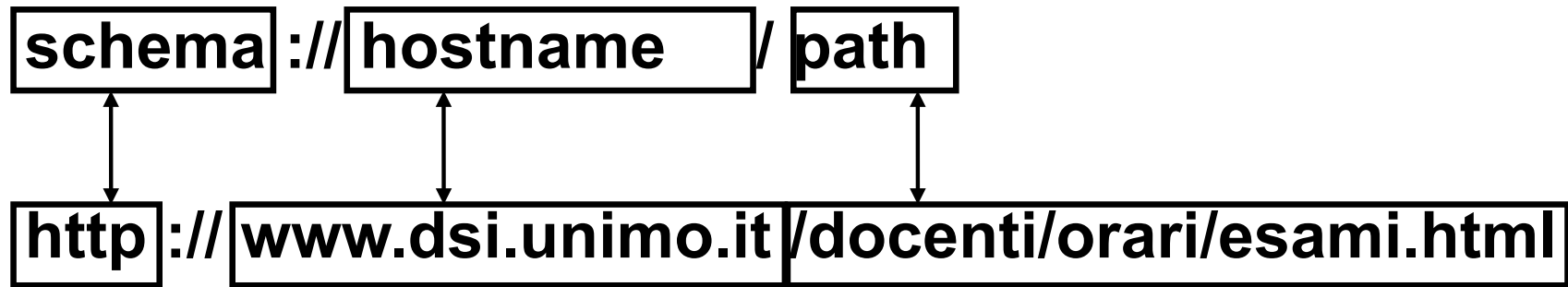
Ingredienti del Web

- **Meccanismi di comunicazione e naming di Internet**
 - Protocollo TCP/IP
 - Sistema DNS
- **Sistema client-server**
- **Integrazione di alcuni standard**
 - **URL: Sistema di naming delle risorse**
 - **HTML: Linguaggio di markup ipertestuale**
 - **HTTP: Protocollo per le richieste risorse**
 - **MIME**

Uniform Resource Locator (URL)

Il sistema di indirizzamento delle risorse è basato su ***Uniform Resource Locator (URL)***, un meccanismo standard per fare riferimento **a tutte le risorse** presenti nel Web, astraendo il reale oggetto delle richieste (file statici, funzionalità, altro)

Campi dell'URL (semplificato)



- **schema**: indica il protocollo impiegato per inviare la risorsa, nel Web HTTP (oggi quasi deprecato, protocollo non sicuro) o HTTPS (versione “sicura” di HTTP, maggiori dettagli in prossima lezione)
- **hostname**: è l'hostname del nodo nel quale risiede la risorsa Web.
- **path**: identifica la risorsa presso il server Web.

Ingredienti del Web

- **Meccanismi di comunicazione e naming di Internet**
 - Protocollo TCP/IP
 - Sistema DNS
- **Sistema client-server**
- **Integrazione di alcuni standard**
 - **URL**: Sistema di indirizzamento delle risorse
 - **HTTP**: Protocollo per le richieste di risorse
 - **HTML**: Linguaggio di markup ipertestuale

Protocollo HTTP

- *HyperText Trasmission Protocol* (HTTP) è il protocollo che permette il reperimento delle risorse Web
- È un protocollo applicativo di tipo **request/reply** basato sulla suite di protocolli TCP/IP

Protocollo HTTP (2)

HTTP usa TCP come protocollo di trasporto

- Il TCP offre un servizio di trasferimento affidabile: i messaggi di richiesta/risposta sono consegnati integri al processo destinatario
- Il client inizia la connessione TCP verso il server sulla porta 80
- Il server accetta la connessione TCP dal client
- I messaggi HTTP di tipo testuale sono scambiati tra browser e Web server
- Alla fine, vi è la chiusura della connessione TCP

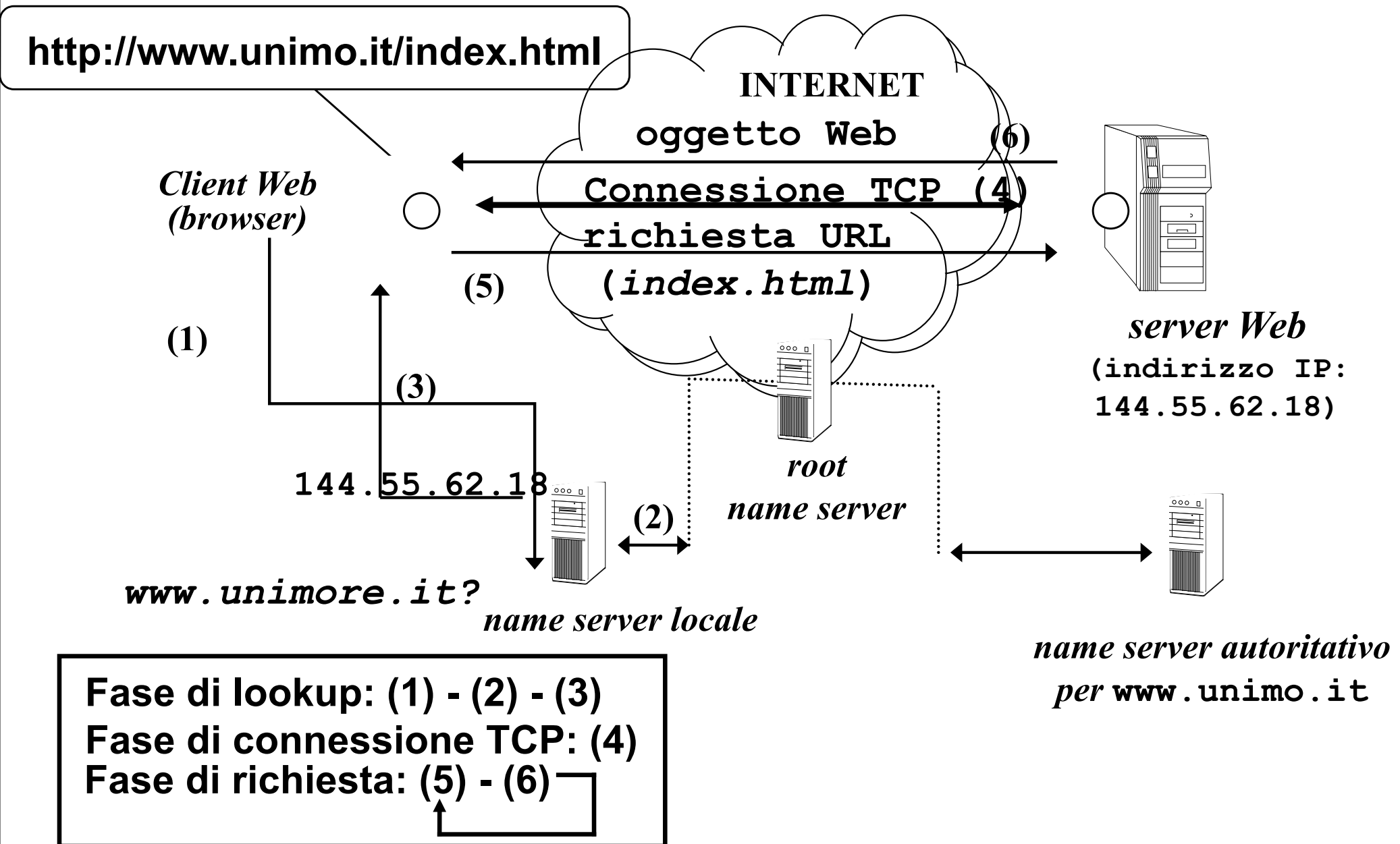
Protocollo HTTP (3)

HTTP è un protocollo stateless (senza stato)

- Il server non conserva alcuna informazione riguardante le richieste dei client passati
- I protocolli che conservano lo stato sono complessi
 - ♦ La storia passata (lo stato) deve essere memorizzata
 - ♦ Se il server/client subiscono un crash, la vista dello stato può essere inconsistente e deve essere ristabilita

Tuttavia, si recupera il concetto di stato e di sessione mediante l'uso dei cookies

Richiesta di una risorsa con HTTP



Ingredienti del Web

- **Meccanismi di comunicazione e naming di Internet**
 - Protocollo TCP/IP
 - Sistema DNS
- **Sistema client-server**
- **“Solo” tre nuovi standard**
 - **URL: Sistema di indirizzamento delle risorse**
 - **HTTP: Protocollo per le richieste risorse**
 - **HTML: Linguaggio di markup ipertestuale**

Istruzioni HTML (extra)

- La descrizione dei contenuti dell'ipertesto viene effettuata inserendo all'interno del testo stesso alcune istruzioni dette *marcatori* o *markup* o *tag* che producono le visualizzazioni e le azioni specificate.
- Le istruzioni HTML sono racchiuse tra parentesi angolari, nella forma **<tag>**, e vengono terminate da un tag di chiusura nella forma **</tag>**. Es.

Schema semplificato di un documento HTML (extra)

```
<HTML>  
  <HEAD>  
    <TITLE>  
      Titolo del documento  
    </TITLE>  
  </HEAD>  
  <BODY>  
    Corpo del documento  
  </BODY>  
</HTML>
```

Immagini in un documento HTML (extra)

- Ciascuna immagine è contenuta in un file differente dal testo.
- Uso di espliciti tag per le immagini:

```
<IMG SRC "Pathname_del_file" >
```

- Possibilità di allineare le immagini al testo:

```
<IMG SRC "Pathname_del_file" align=middle>
```

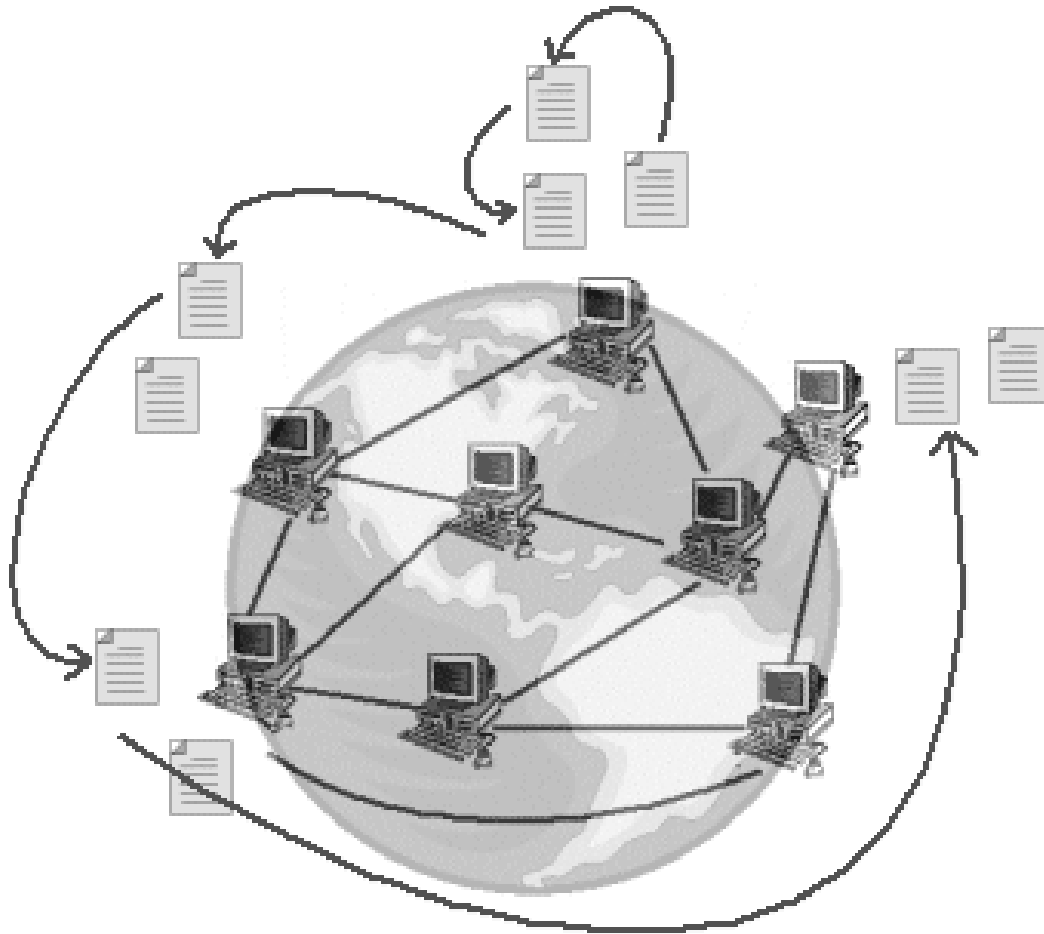
Tag àncora (extra)

- L'istruzione più innovativa dell'HTML è l'**àncora** delimitata dai tag **<A>...**, in quanto tale elemento permette di trasformare un normale testo in ipertesto multimediale.
- Un àncora può far riferimento ad una sezione della stessa pagina oppure **ad una qualsiasi risorsa** (testuale, multimediale, eseguibile) presente sul Web, denotata mediante un **URL** che va inserito all'interno del tag àncora.

```
<A  
HREF="http://www.unimo.it/studenti/erasmus.html">  
Programma Erasmus </A>
```

- Il testo **Programma Erasmus** viene visualizzato in modo differente e risulta un link simbolico selezionabile via mouse.

Il tag à ancora consente l'ipermedialità su scala geografica



Conseguenza dirompente dal punto di vista storico e sociale: l'informazione non è più organizzata in rigidi schemi (es., biblioteca), ma è accessibile (anche) mediante “link” decisi da singole persone, aziende, organizzazioni, ...

Ogni link inserito tra una pagina Web e un'altra determina una nuova modalità di accesso all'informazione

WWW- Lato *Client HTTP*

Browser

- **Il browser è un'applicazione software che svolge il ruolo di interfaccia fra l'utente e il WWW, mascherando la complessità di Internet**
- **Diventa un client Web per recuperare informazioni dai server Web**
- **Funzioni principali**
 - instaura una connessione TCP con il server tramite cui invia opportuni messaggi al server Web per ottenere le risorse richieste
 - interpreta il codice ipertestuale HTML
 - elabora il codice allo scopo di visualizzare in modo appropriato il contenuto delle pagine sullo schermo

Azioni di un browser (2)

Fase di lookup

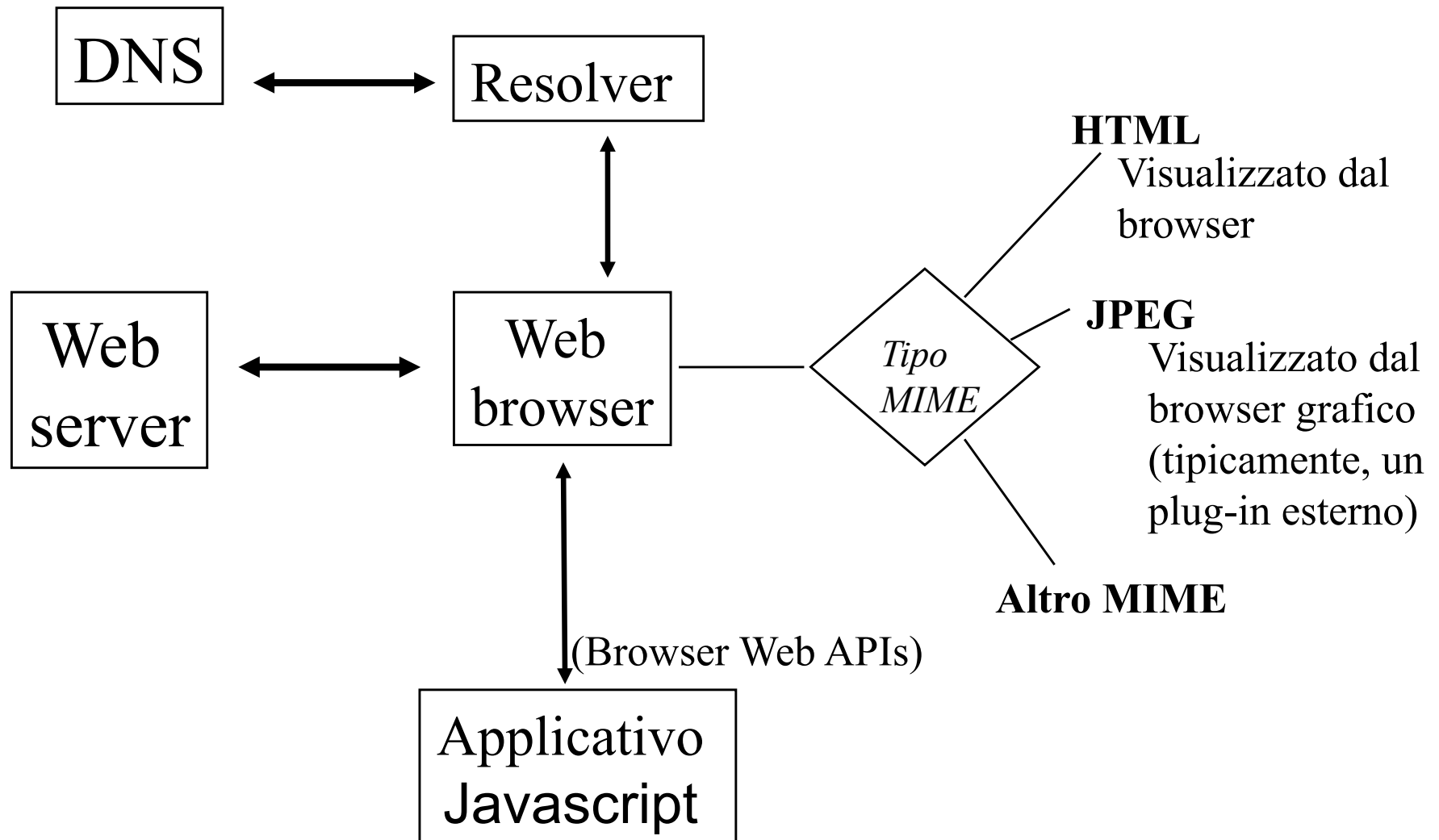
- Il browser acquisisce la componente *hostname* dall'URL
- Il browser invoca il *resolver* per conoscere, tramite il DNS, l'indirizzo IP corrispondente all'*hostname*
- Se esistente, il DNS restituisce l'indirizzo IP

Azioni di un browser (3)

Fase di richiesta

- Una volta individuato l'indirizzo IP, il browser stabilisce una connessione TCP con l'host avente l'indirizzo IP individuato
- Sfruttando questa connessione, il browser richiede mediante il protocollo HTTP la risorsa specificata nell'URL tramite una richiesta HTTP
 - L'hostname inserito e impiegato per la risoluzione DNS viene anche utilizzato per essere inserito come header 'Host' per la funzionalità di virtual hosting di HTTP sul Web server

Software in esecuzione sulla macchina client



Caching nel browser (extra)

- Il browser gestisce uno spazio disco (di dimensione predefinita) in cui memorizza i file dei server Web recuperati corrispondenti a:
 - pagine HTML
 - immagini
- In realtà, prima di effettuare una richiesta al server Web, il browser controlla se nella cache vi è una copia dell'oggetto richiesto