

CONCETTI INTRODUTTIVI E TERMINOLOGIA

Sommario

- I tre grafi della rete
- Caratteristiche fondamentali del Web
- Terminologia
- Architettura
- Web server

I Tre Grafi della Rete (*)

- Da un punto di vista strutturale la rete può essere modellata secondo tre diversi grafi
 - Il grafo (albero) dei nomi di dominio (Domain Name System – DNS)
 - Il grafo del Web
 - Il grafo delle connessioni fisiche
- Il corso si concentra **principalmente** (ma **non esclusivamente**) sul grafo del Web

(*) dall'omonimo capitolo di [Luccio, Pagli, 2007]

Il grafo del DNS (1)

- Ogni risorsa sul web è identificata univocamente da un indirizzo simbolico
 - Ad esempio il sito del **Dipartimento di Informatica** dell'**Università di Bari** in **Italia** si trova all'indirizzo
`http://www.di.uniba.it`
- Tralasciando "http://www.", **di.uniba.it** indica il nome del relativo dominio

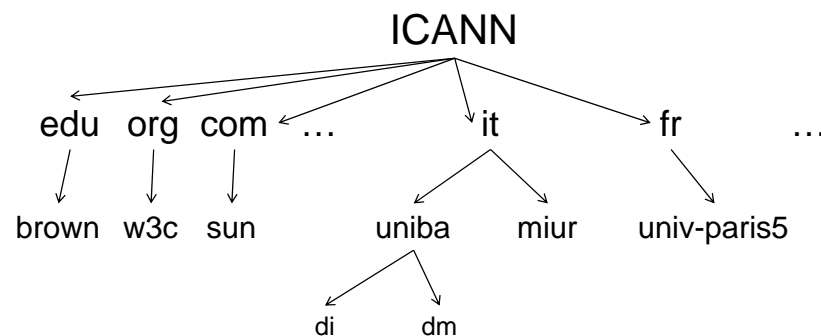
Il grafo del DNS (2)

- I nomi di dominio sono definiti dall'**Internet Corporation for Assigned Names and Numbers – ICANN**
 - Organizzazione del Dipartimento del Commercio degli USA
 - Definisce l'identificatore delle macchine (logiche o fisiche) su cui si trovano le risorse web

Il grafo del DNS (3)

- L'ultimo nome di ogni dominio identifica
 - il tipo di organizzazione in cui risiede l'host **OPPURE**
 - il paese in cui si trova
- Esempi
 - di.uniba.it perché si trova in **Italia**
 - math-info.univ-paris5.fr perché si trova in **Francia**
 - cs.brown.edu perché è un'organizzazione **educational**
 - sun.com perché è un'organizzazione **commercial**
 - un.gov perché è un'organizzazione di **government**
 - w3c.org perché è un'organizzazione non governativa
 - sourceforge.net perché è un **network**

Il grafo del DNS (4)



- È un **albero!**

Il grafo del Web (1)

- È l'insieme di tutte le risorse accessibili sul Web, dove
 - Ogni risorsa è un nodo del grafo
 - I link che da una risorsa permettono di passare a un'altra sono gli archi del grafo
 - I link rappresentano **connessioni logiche**

Il grafo del Web (2)

- È un **grafo** con elevata dinamicità
 - Nuove risorse sono continuamente aggiunte
 - Risorse vecchie vengono eliminate
 - Si stabiliscono continuamente nuovi archi
- È difficile
 - censire le risorse
 - stabilire la forma
 - definire algoritmi efficaci ed efficienti per l'attraversamento

Il grafo delle Connessioni Fisiche

- Ogni computer che vuole accedere a una risorsa Internet deve essere connesso ad un Internet Service Provider – ISP
- Per mezzo di specifici apparati ogni ISP è in grado di accedere al contenuto di altri
- Esistono diversi tipi di ISP: istituzionali, commerciali, no-profit, ...
- I dispositivi (fisici e logici) che permettono il collegamento tra diversi ISP sono

argomento di altro corso

Web: Caratteristiche Fondamentali

- Collezione di **risorse** fisicamente sparse nel mondo (**Hypermedia distributed system**), su macchine diverse,
 - accessibili attraverso un particolare sistema di indirizzamento (faremo riferimento a **URI**),
 - mediante opportuni protocolli (principalmente **HTTP**, ma non solo),
 - sfruttando opportuni linguaggi (principalmente **XHTML**, ma non solo)

Sistema ipermediale distribuito

- Le risorse del sistema
 - sono costituite da documenti che si poggiano su media di varia natura e opportunamente collegati tra di loro
 - sono distribuite su diverse macchine opportunamente interconnesse

Indirizzi Internet (1)

- Dal punto di vista di un utente umano Internet può essere percepita come un insieme di *nodi*, ciascuno dei quali è identificato da un nome
 - Ad es. la pagina web del Dipartimento di Informatica ha nome <http://www.di.uniba.it>
- Per un computer i nodi sono identificati da un indirizzo numerico
 - Per la pagina web del Dipartimento di Informatica è 193.204.187.130

Indirizzi Internet (2)

- L'indirizzo numerico è l'*Internet Protocol Address* della macchina connessa ad Internet che contiene quella pagina
- L'Internet Protocol Address è sempre un numero a 32 bit, espresso come 4 numeri a 8 bit, separati dal simbolo .

Indirizzi Internet (3)

- L'IP address è gestito facilmente da un computer, ma di difficile memorizzazione da parte degli utenti umani
- Il nome di dominio permette facile gestione per gli umani ma deve essere tradotto in IP address per la gestione automatica

URL (1)

- *Uniform* (o *Universal*) *Resource Locator*, è usato per identificare univocamente un *documento* (una *risorsa*) su Internet
- Esistono molteplici risorse, quindi esistono molteplici formati per gli URL

URL (2)

- Lo schema generale è
scheme:object-address
- Scheme è il protocollo di comunicazione
 - Ad es: http, ftp, gopher, telnet, file, mailto, news
 - In questo corso ci concentreremo sullo scheme http, usato per richiedere/inviare documenti XHTML
 - Talvolta useremo lo scheme file

URL (3)

- L'object-address dipende dallo scheme
 - Per http l'object-address è della forma
//fully-qualified-domain-name/path-to-doc
 - Per file l'object-address è della forma
//path-to-doc
- path-to-doc indica il cammino per raggiungere il documento all'interno dell'host, analogamente a quanto avviene nel file system

URL (4)

- L'URL può essere una qualsiasi stringa alfanumerica in cui non compaiono i caratteri **blank, ,, ;, &**
 - Alcuni browser ammettono l'uso di tali simboli, ma è un comportamento fuori standard
- Qualora fosse necessario inserire tali caratteri nell'URL, è necessario usare il simbolo % seguito dal codice ASCII in esadecimale del carattere
 - Ad es. la stringa "Programmazione Web" diventa "Programmazione%20Web"

Porte (1)

- L'hostname è il nome del computer che
 - memorizza il documento oppure
 - fornisce accesso al documento memorizzato su altro computer
- I messaggi a una macchina host devono essere diretti ad un apposito processo eseguito sull'host stesso
 - Ognuno di questi processi è identificato da un **port number**

Porte (2)

- La porta di default a cui faremo riferimento è la porta 80
 - Se il server è stato configurato per usare un'altra porta **pp**, questa deve essere specificata facendo seguire alla URL la stringa

:pp

MIME (1)

- Multipurpose Internet Mail Extensions
- Originariamente è stato sviluppato per definire la forma dei documenti inviati attraverso la mail
- Più in generale oggi esprime la forma dei documenti che un browser riceve da un server web

MIME (2)

- Quando un browser riceve dal server web un doc, legge la specifica di formato MIME per capire come trattare il doc
 - Testo
 - Immagini
 - Video
- Le specifiche MIME hanno la forma
type/subtype

MIME (3)

- I tipi più comuni sono:
 - **text**, con subtype **plain** e **html**
 - **image**, con subtype **gif** e **jpeg**
 - **video**, con subtype **mpeg** e **quicktime**

Architettura (1)

- Un utile modello per rappresentare lo scambio di informazioni sul Web è quello **client-server**:
 - Il client **richiede un servizio**
 - Il server dopo aver verificato i diritti di accesso al servizio, **lo fornisce** al client

Architettura (2)

- Più precisamente:
 - L'utente specifica la URL che identifica la risorsa cercata
 - Il browser (**web client**) mediante il protocollo HTTP invia una richiesta al web server
 - Il server elabora la richiesta e mediante il protocollo HTTP invia la risposta al browser
 - La risposta è in forma di file scritto in qualche linguaggio che può essere interpretato dal browser
 - Il browser legge e interpreta il file e mostra il risultato dell'elaborazione all'utente

HTTP

- HyperText Transfer Protocol
- È il protocollo usato per le comunicazioni sul web
- Prevede due fasi diverse: **request** e **response**
- Entrambe le fasi comprendono a loro volta due parti:
 - **header** che contiene informazioni riguardo le modalità di comunicazione
 - **body** che contiene i dati della comunicazione

Web Server

- Strato di software posto su una macchina ad hoc che risponde alle richieste di un client, fornendo le risorse richieste
- URI (Uniform Resource Identifier)
- Il server e il client Web comunicano mediante il protocollo Hypertext Transfer Protocol (HTTP)
 - Platform-independent

Architettura Sistema (1)

- Multi-tier application (*n*-tier application)
 - Livello Information (livello dati)
 - È il livello inferiore
 - Gestisce i dati dell'applicazione in un apposito DB, in genere basato su un RDBMS
 - Livello Logic
 - È il livello intermedio
 - Implementa la logica dell'applicazione e la sua presentazione
 - Controlla le interazioni tra i client dell'applicazione e i dati dell'applicazione

Architettura Sistema (2)

- Livello Client
 - È il livello superiore
 - Realizza l'interfaccia dell'applicazione
 - Gli utenti interagiscono direttamente con l'applicazione attraverso questo livello

Script lato Client e Script lato Server (1)

- Script lato Client
 - Validano l'input dell'utente
 - Minimizzano le richieste che devono essere inviate al server
 - Permettono l'accesso al browser
 - Migliorano le pagine web mediante le tecnologie DHTML, controlli ActiveX, applet, ...

Script lato Client e Script lato Server (2)

- Script lato Server
 - Vengono eseguiti sul server
 - Generano risposte specifiche per ogni client
 - Ampio spettro di capacità di programmazione
 - Permettono l'accesso al software lato server che estende le funzionalità del server stesso

Web Resources

- www.microsoft.com/msdownload/ntoptipack/askwis.asp
- www.w3.org/Protocols
- www.apache.org
- httpd.apache.org
- httpd.apache.org/docs-2.0
- www.apacheweek.com
- linuxtoday.com/stories/18780.html
- www.iianswers.com
- www.iiadministrator.com