UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

STRUMENTI E APPLICAZIONI DEL WEB

Il Web: origini e caratteristiche

Marco Viviani

Anno Accademico 2025-2026

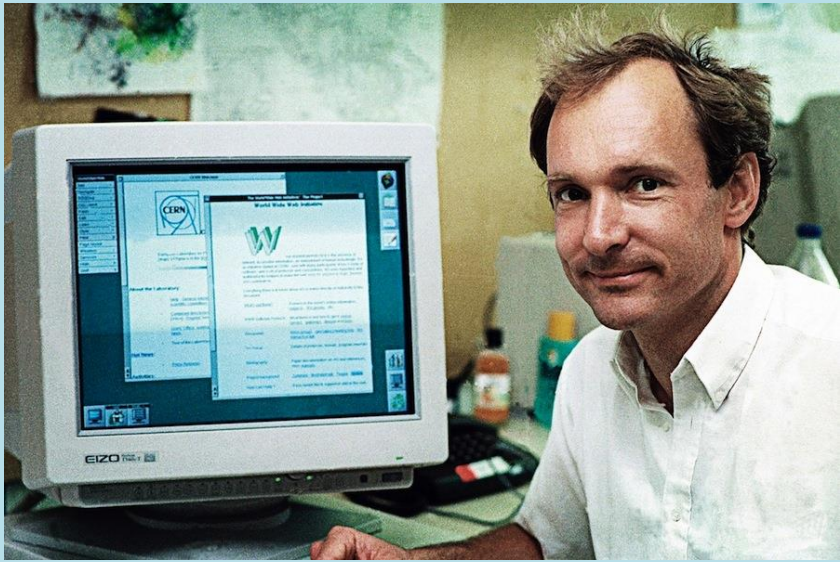
Introduzione al Web e alle sue principali caratteristiche

Il World Wide Web (1/6)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Il **World Wide Web (WWW)** rappresenta uno dei principali servizi di Internet.

- La concezione del World Wide Web si deve a **Tim Berners-Lee**, elaborando un un progetto precedente (con Robert **Robert Cailliau**) di condivisione di documentazione scientifica in formato elettronico indipendentemente dalla piattaforma.

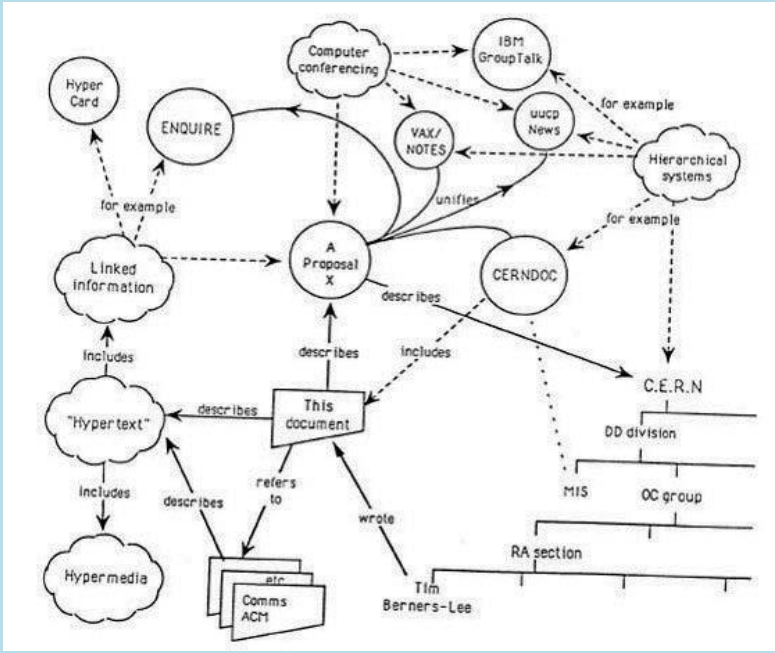


Il World Wide Web (2/6)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- La proposta fu ideata per fornire un sistema di comunicazione più efficace all'interno dell'Organizzazione europea per la ricerca nucleare (**CERN: Conseil Européen pour la Recherche Nucléaire**), ma Berners-Lee si rese presto conto delle potenzialità globali dell'idea.

Il logo storico del WWW disegnato da Robert Cailliau nel 1990



L'architettura del WWW nella proposta originale di Berners-Lee

Il World Wide Web (3/6)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Il Web permette di navigare e usufruire di un insieme molto vasto di **contenuti amatoriali e professionali (multimediali e non)** collegati tra loro e di ulteriori servizi accessibili a tutti o ad una parte selezionata degli utenti di Internet.

- I concetti fondamentali alla base del Web sono quello di **ipertesto (*hypertext*)** e **collegamento ipertestuale (*hyperlink*)**.

    - L'ipertesto è un testo strutturato che utilizza collegamenti logici tra i pagine che contengono testo (e altro contenuto multimediale).

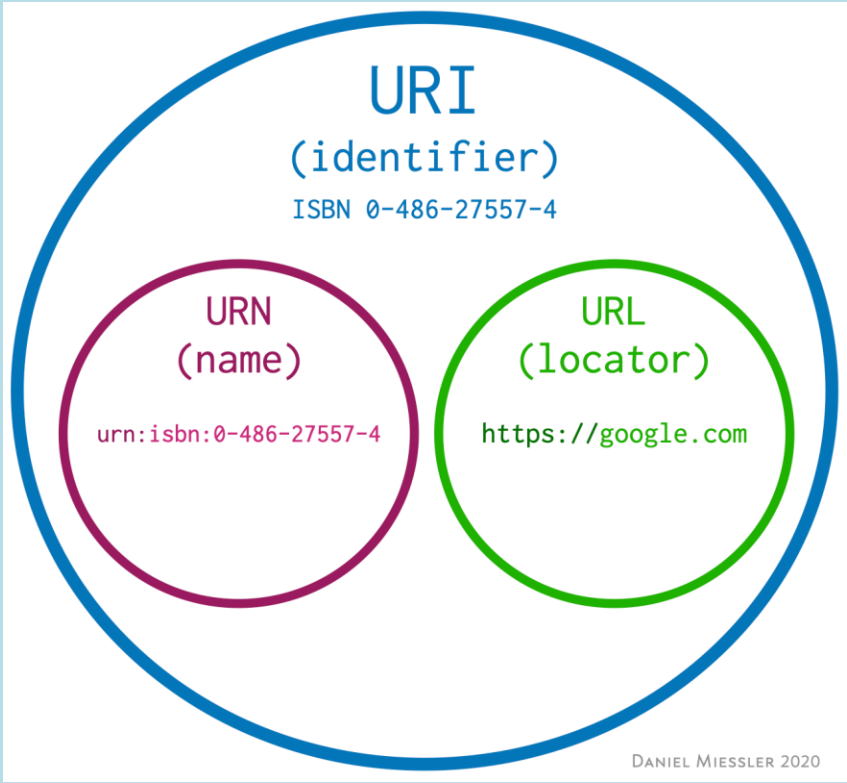
Il World Wide Web (4/6)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Identificare le risorse **univocamente** sulla rete (e non…).
- **Uniform Resource Identifier (URI)**: sequenza di caratteri che identifica univocamente una risorsa generica.
  - Sono **esempi di URI**: un indirizzo Web (**Uniform Resource Locator, URL**), un documento, un'immagine, un file, un indirizzo di posta elettronica, ecc.

Differenze: URI, URL, URN

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



https://danielmiessler.com/study/difference-between-uri-url/

Il World Wide Web (5/6)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- **Linguaggi di markup**: Berners-Lee ha sviluppato il linguaggio di pubblicazione **HyperText Markup Language (HTML)**, attraverso il quale è possibile indicare nel testo:
  - La struttura;
  - Il significato semantico;
  - La modalità di presentazione dei contenuti.
- Un documento in HTML costituisce un **dato semi-strutturato**.

Una parentesi: dati strutturati

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

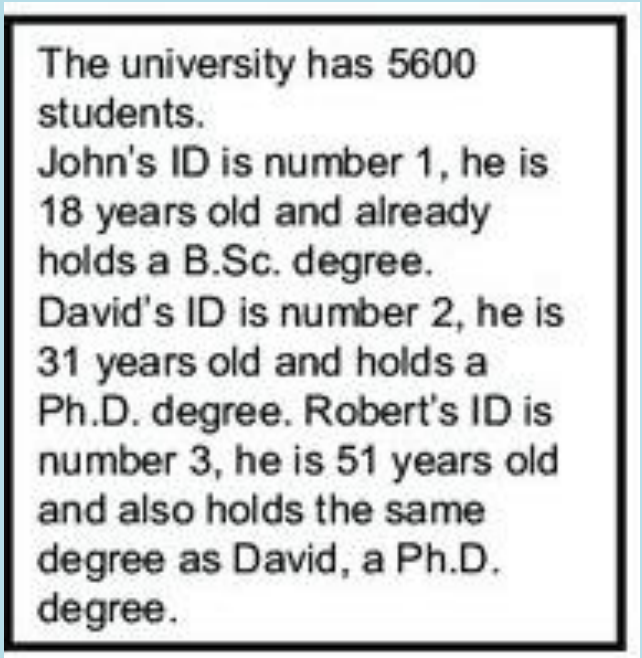
- Dati memorizzati in **database**, organizzati secondo schemi e tabelle rigide.
- È il tipo di dati più adatto ai modelli di **gestione relazionale** delle informazioni.
- La gestione dei database è affidata ad un **Database Management System (DBMS)**.

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

Una parentesi: dati non strutturati

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

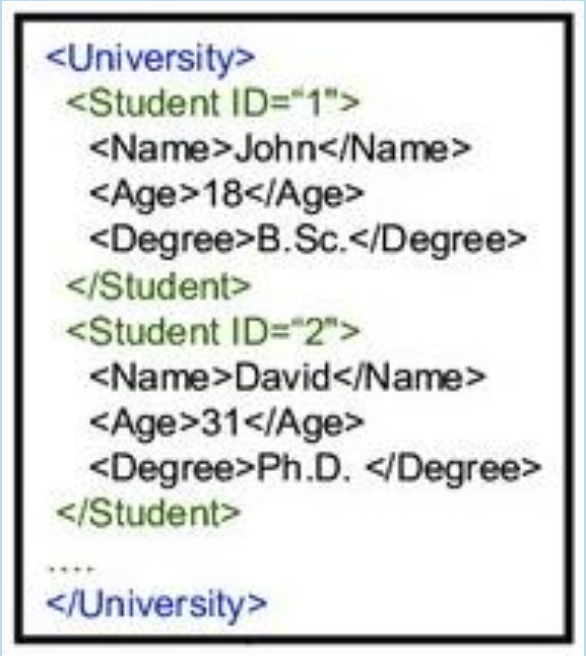
- Dati memorizzati **senza alcuno schema**.
- Un esempio potrebbe essere costituito da **testi narrativi** prodotti mediante uno dei più diffusi software di editing testuale.
- In questo caso, i sistemi di gestione dei dati che possono essere utilizzati sono quelli basati su **modelli di Information Retriveval (IR)**.


The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Una parentesi: dati semi-strutturati

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Una forma di dati strutturati che **non obbedisce alla struttura tabellare** dei modelli di dati associati ai database relazionali o ad altre forme di tabelle di dati.
- Tuttavia, contiene *tag* o altri **marcatori** per separare gli elementi semantici e rafforzare le gerarchie di record e campi all'interno dei dati.
- Per questo motivo, è nota anche come **struttura autodescrittiva**.



Linguaggi di markup procedurali e dichiarativi

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- I **linguaggi di markup di tipo procedurale** indicano le procedure di trattamento del testo aggiungendo le istruzioni che devono essere eseguite per visualizzare la porzione di testo referenziata.
  - troff, LaTex, ...
- I **linguaggi di markup di tipo dichiarativo** lasciano la scelta del tipo di rappresentazione da applicare al testo al software che di volta in volta lo riprodurrà.
  - risultano più vantaggiosi perché si concentrano sui problemi strutturali di leggibilità e prescindono in fase di lettura dal software con cui sono stati generati. Sono, in altre parole, quelli che permettono di garantire una corretta separazione tra struttura e visualizzazione;
  - XML, HTML, XHTML.

Basi di XML (1/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- XML (*eXtensible Markup Language*, lett. "linguaggio di marcatura estendibile") è un metalinguaggio per la definizione di linguaggi di markup.

- È un linguaggio basato su un **meccanismo sintattico** che consente di definire e controllare il significato degli elementi contenuti in un documento o in un testo.

- Le tecnologie XML sono standardizzate dal consorzio **W3C (*World Wide Web Consortium*)**, un'organizzazione non governativa internazionale che ha come scopo quello di sviluppare tutte le potenzialità del World Wide Web.
    - Approfondimento nelle prossime lezioni.

Basi di XML (2/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- L'unità base di ogni documento XML sono gli **elementi**.
- Un elemento corrisponde a uno specifico blocco di testo all'interno di un documento, marcato attraverso una **coppia di tag**.
  - Start tag, End tag
  - <paragrafo>Testo del paragrafo</paragrafo>
- Un **elemento** è perciò costituito da una coppia di tag e dal suo contenuto.

Nomi di elementi

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Elementi diversi devono essere contraddistinti da nomi diversi.
- Il nome di un elemento viene definito **Generic Identifier (GI)**.
  - XML è «case sensitive»
  - <paragrafo> è diverso da <PARAGRAFO>

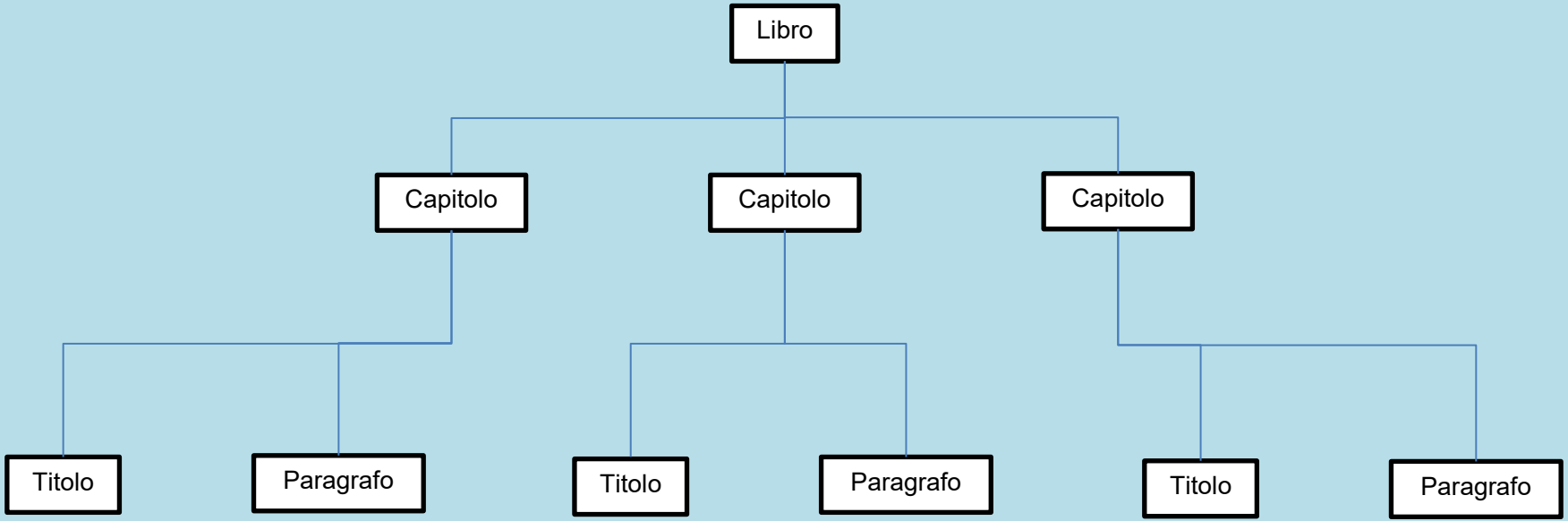
Struttura di un documento XML

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Un **documento XML** è intrinsecamente caratterizzato da una **struttura gerarchica**.
- Esso è composto da. **diversi elementi.**
  - Ciascun elemento rappresenta un componente logico del documento e può contenere altri elementi (sotto-elementi) o del testo.
- L'organizzazione degli elementi segue un **ordine gerarchico o arboreo** che prevede un elemento principale, chiamato ***root element*** o semplicemente ***root*** o ***radice***.
- La radice contiene l'insieme degli altri elementi del documento.
  - Possiamo rappresentare graficamente la struttura di un documento XML tramite un **albero**, generalmente noto come ***document tree***.

Un esempio di struttura di albero XML (1/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



Un esempio di struttura di albero XML (2/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

```
<libro>
        <capitolo>
                <titolo>Titolo del capitolo</titolo>
                <paragrafo>Paragrafo 1</paragrafo>
                <paragrafo>Paragrafo 2</paragrafo>
        </capitolo>
        <capitolo>
                …
        </capitolo>
        <capitolo>
                …
        </capitolo>
</libro>
```

Esercizio

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Generare un file XML da zero, rispetto ad una realtà che possa essere espressa in forma gerarchica, in cui:
  - Ci sia un elemento radice
  - La radice abbia quattro elementi «figli»
  - Ogni elemento figlio della radice abbia a sua volta quattro elementi «figli»

Altri elementi di XML

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- A ogni elemento può essere associato un insieme di **attributi**, che possono essere utilizzati per specificarne con più dettaglio le proprietà.
- Gli attributi possono essere dichiarati solo in uno start tag.
  - <capitolo numero = "1" >Testo del capitolo</capitolo>
- Gli attibuti possono SEMPRE essere sostituiti con dei tag (scelta consigliata)

```
<capitolo>
    <numero>1</numero>
    <testo>Testo del capitolo</testo>
</capitolo>
```

Esercizio

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Pensare a un caso di un elemento con attributo e a come trasformarlo mediante l'utilizzo di altri tag.

Un esempio dell'uso dei commenti

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Un commento è una particolare sequenza che può essere inserita nel codice XML per fornire informazioni sul codice stesso o per evitare che parte del codice sia elaborato.
  - <!-- testo del commento-->

```
<!-- inizio del libro-->
<capitolo>
    <titolo>Titolo</titolo>
    <paragrafo>Primo paragrafo</paragrafo>
    <paragrafo>Secondo paragrafo</paragrafo>
    <!-- <paragrafo>Terzo paragrafo</paragrafo> -->
</capitolo>
```

File XML

- Un **file XML** è un documento di testo, con estensione .xml. È diviso in **due parti principali**:
  - **Prologo**, dove si specificano:
    - La versione del linguaggio e opzionalmente il set di codifica dei caratteri usato nel documento.
      - <? xml version = "1" encoding = "UTF-8" ?>
    - Istruzioni di elaborazione, per il collegamento ai fogli di stile.
      - <?xml-stylesheet type = "text/css" href= "stile.css" ?>
    - Una Doctype Declaration, obbligatoria se il documento è associato a una Document Type Definition.
      - <! DOCTYPE grammatica SYSTEM "text/css" [] >
  - L'**istanza del documento**.

Document Type Definition

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Il *Document Type Definition* (definizione del tipo di documento) è uno strumento utilizzato dai programmatori il cui scopo è quello di definire le componenti ammesse nella costruzione di un documento XML.
  - Definisce gli **elementi leciti** all'interno del documento. Non si possono usare altri elementi se non quelli definiti. Una specie di "vocabolario" per i file che lo useranno.
  - Definisce la **struttura di ogni elemento** (Es., cosa può contenere ciascun elemento, l'ordine, la quantità di elementi che possono comparire, ecc.). Una specie di "grammatica".
  - Dichiara una serie di **attributi per ogni elemento** e che valori possono o devono assumere questi attributi.
  - Fornisce infine alcuni **meccanismi per semplificare la gestione del documento** (Es., la possibilità di importare parti di altri DTD).

Un semplice esempio di DTD

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Il seguente DTD:

```
<! ELEMENT persona (nome, cognome)>
<! ELEMENT nome (#PCDATA) > (PCDATA → "Parsed Character Data")
<! ELEMENT cognome (#PCDATA) >
```

definisce una struttura così composta:

```
<persona>
        <nome>Mario</nome>
        <cognome>Rossi</cognome>
</persona>
```

Metadati associati (1/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Un **metadato** (dal greco μετὰ "oltre, dopo, per mezzo" e dal latino datum "ciò che è dato " - plurale: data), letteralmente "(dato) per mezzo di un (altro) dato", è un'informazione che descrive un insieme di dati.

- Il *Dublin Core* (DC) è un sistema di metadati costituito da un nucleo di elementi essenziali ai fini della descrizione di qualsiasi materiale digitale accessibile via rete informatica.

http://dublincore.org/documents/1999/07/02/dces/

Metadati associati (2/2)

- In XML, si tratta di tag non utilizzati come marcatori, ma unicamente per descrivere le proprietà del testo
  - <meta name = "DC.title" lang = "it" content = "I promessi sposi" />
  - <meta name = "DC.creator" content = "Alessandro Manzoni" />
  - <meta name = "DC.subject" content = "Letteratura italiana" />
  - <meta name = "DC.publisher" content = "Mondadori" />
  - <meta name = "DC.type" content = "Text" />

HTML

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- HTML è l'acronimo di *HyperText Markup Language* ("Linguaggio di contrassegno per gli Ipertesti"). Si tratta di un linguaggio di markup che permette di indicare come disporre (e come dovranno apparire) gli elementi all'interno di una pagina Web.

- Con HTML quindi indichiamo, attraverso i *tag*, quali elementi dovranno apparire su uno schermo e come essi debbano essere disposti. Tutte queste indicazioni sono contenute in un documento HTML, spesso detto «Pagina HTML».

- Una pagina HTML è rappresentata da un file di testo, ovvero un file che possiamo modificare con programmi come «blocco note» e in genere hanno un nome che finisce con l'estensione .html.

La struttura ad albero del documento HTML (1/3)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

```
<html>
  <head>
    <title>Page title</title>
  </head>
  <body>
    <h1>This is a heading</h1>
    <p>This is a paragraph.</p>
    <p>This is another paragraph.</p>
  </body>
</html>
```

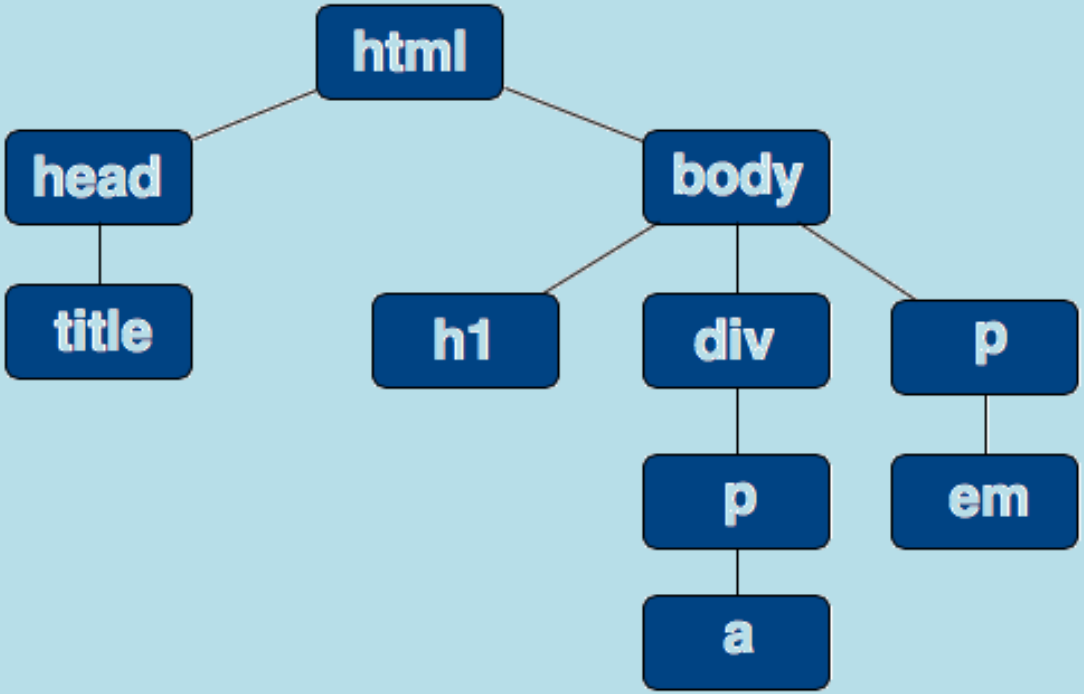
La struttura ad albero del documento HTML (2/3)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

```
<html>
  <head>
    <title>Page title</title>
  </head>
  <body>
    <h1>This is a heading</h1>
    <p>This is a paragraph.</p>
    <p>This is another paragraph.</p>
  </body>
</html>
```

La struttura ad albero del documento HTML (3/3)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



Gli attributi in HTML

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- I tag HTML possono essere corredati di uno o più **attributi**, che servono per meglio specificare la funzione o la tipologia dell'elemento, per memorizzare dati o per arricchire di significato il contenuto.

- Un tag con attributi si scrive in questo modo:
  - <tag attributo1="valore1" attributo2="valore2">
  - I valori sono tipicamente racchiusi tra virgolette " ", ma è possibile anche utilizzare gli apici ' ';
  - Si scrivono lasciando almeno uno spazio dopo il nome dell'elemento nel tag di apertura (o nell'unico tag nel caso di elementi non contenitori).

Separare il contenuto dal *layout*

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- In passato si utilizzavano alcuni tag HTML per definire font (tipi di carattere), i colori o le dimensioni degli oggetti sullo schermo.
- Oggi il quadro è definitivamente cambiato e molte di quelle funzionalità sono deprecate favorendo una divisione dei compiti più chiara tra diversi strumenti:
  - HTML
    serve a definire quali sono gli elementi in gioco, stabilire collegamenti (link) tra le pagine e l'importanza (non la forma o il colore) che hanno i testi, creare form per gli utenti, fissare titoli, caricare immagini, video, etc.
  - CSS (*Cascading Style Sheets*) o fogli di stile.
    - Si tratta di una serie di regole che permettono di definire l'aspetto (lo stile) che devono assumere gli elementi sulla pagina. Dimensioni, colori, animazioni, ogni caratteristica visuale può essere manipolata.

Il World Wide Web (6/6)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Protocollo di trasferimento (*Transfer Protocol*)

- L'*HyperText Transfer Protocol* (HTTP) (protocollo di trasferimento di un ipertesto) è un protocollo a livello applicativo usato come principale sistema per la trasmissione d'informazioni sul Web.

Il modello Client/Server (1/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- L'**HTTP** è un protocollo che lavora con un'**architettura di tipo Client/Server**: il client esegue una richiesta e il server restituisce la risposta.
- Nell'uso comune il ***client*** corrisponde al *browser* ed il ***server*** la macchina remota su cui risiede il sito Web.
- Vi sono quindi **due tipi di messaggi HTTP**: messaggi richiesta e messaggi risposta.

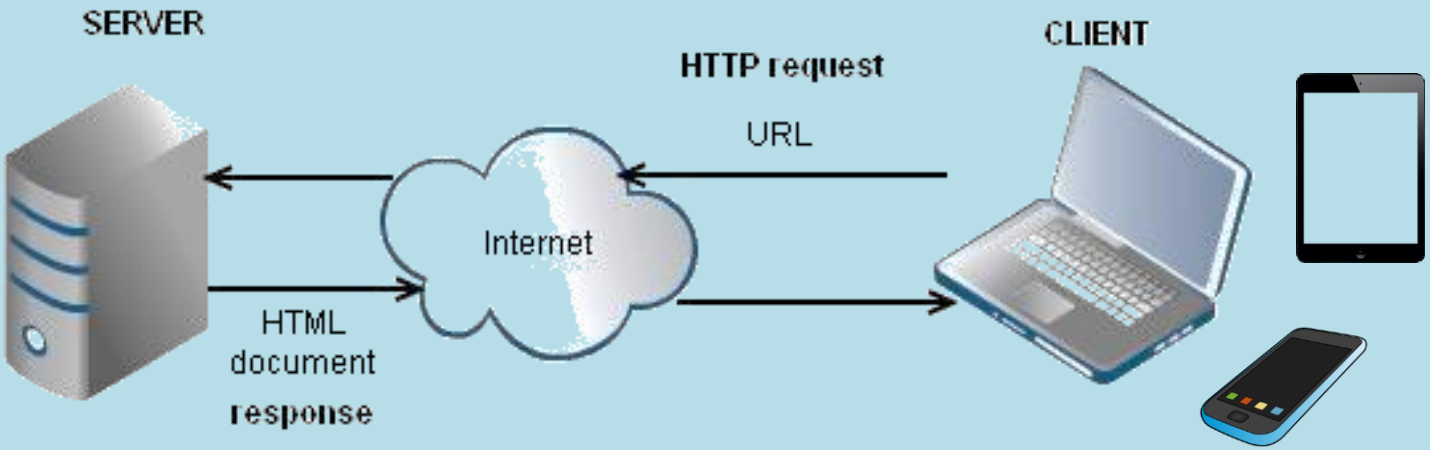
Il *browser*

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Un **Web brow**ser è un'applicazione per l'acquisizione, la presentazione e la navigazione di risorse sul Web.
    - Tra i **browser più utilizzati** vi sono Google Chrome, Internet Explorer, Mozilla Firefox, Microsoft Edge (uscito con Windows 10), Safari, Opera.
- Ci sarà una **lezione dedicata** rispetto ai browser.

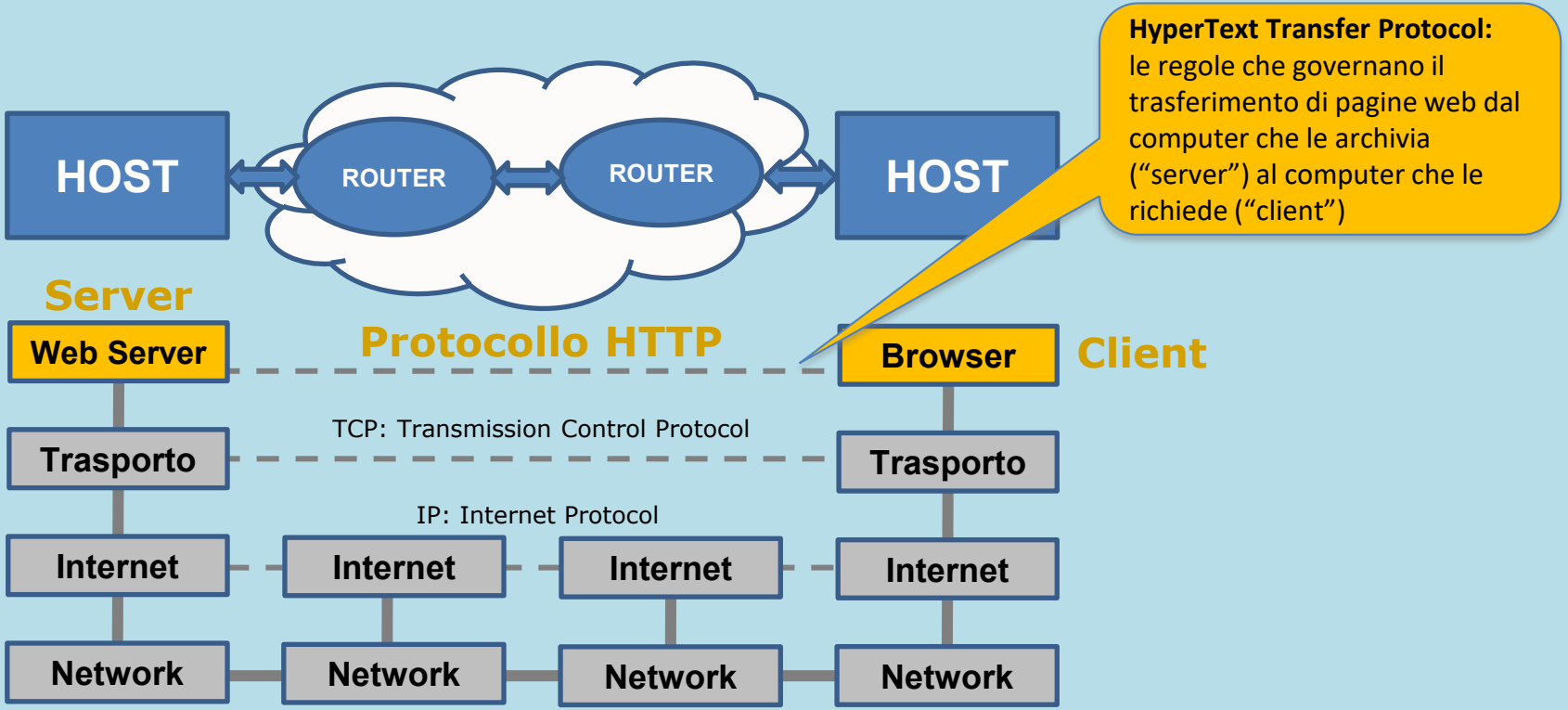
Il modello Client/Server (2/2)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



Il protocollo HTTP

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



I messaggi di errore più comuni

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- **400 Bad Request**. La risorsa richiesta non è comprensibile al server.
- **404 Not Found**. La risorsa richiesta non è stata trovata e non se ne conosce l'ubicazione. Di solito avviene quando l'URI è stato indicato in modo incorretto, oppure è stato rimosso il contenuto dal server.
- **500 Internal Server Error**. Il server non è in grado di rispondere alla richiesta per un suo problema interno.
- **505 HTTP Version Not Supported**. La versione di http non è supportata.

404 «not found»

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

Google

404. That's an error.

The requested URL /doesntexist was not found on this server. That's all we know.

Recap: Internet VS World Wide Web

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- **Internet** è una rete di reti, una interconnessione globale tra reti informatiche di natura e di estensione diversa.
- Tale interconnessione è resa possibile da una suite di protocolli di rete comune chiamata «TCP/IP» (si veda la lezione sulle reti di calcolatori).
- Il **World Wide Web** è una collezione di documenti collegati tra loro tramite link.
- Si basa su URL, protocollo HTTP, HTML e Web browser.
- Il WWW è uno dei servizi supportati da Internet.

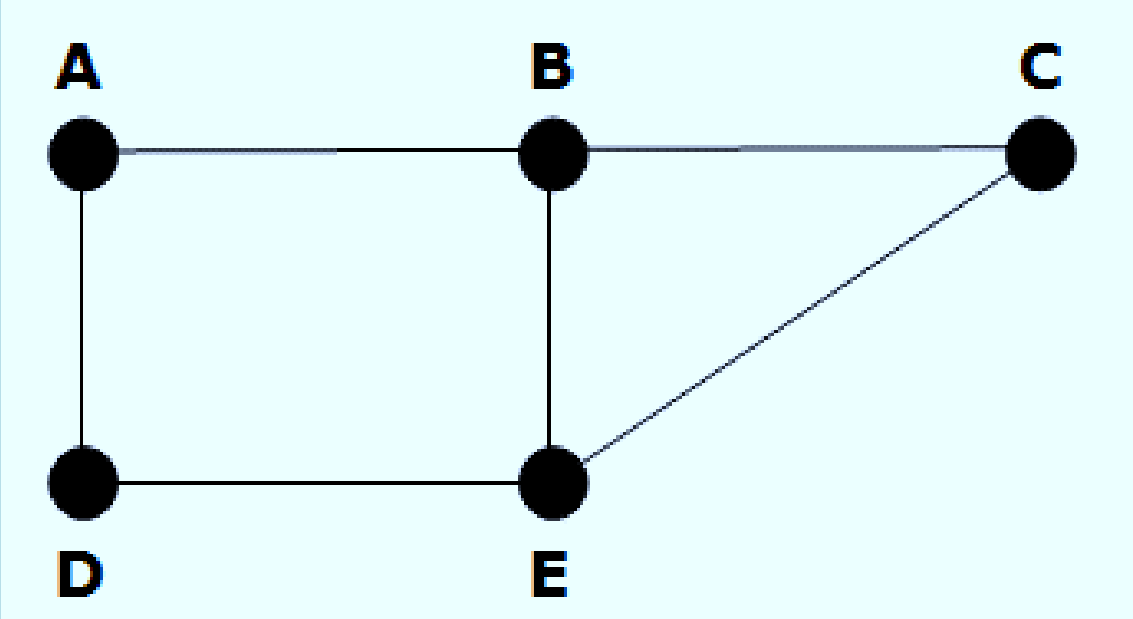
Il grafo del Web

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Si può pensare all'insieme dei documenti presenti sul Web come a un **grafo**, in cui:
  - I nodi sono gli URL;
  - C'è un arco fra il nodo *x* e il nodo *y* quando la pagina che corrisponde all'URL *x* contiene un link verso l'URL *y*.
- Questo grafo è chiamato **grafo del Web**. Ovviamente, si tratta di un grafo dinamico che cambia in continuazione.

Una parentesi… Cos'è un grafo?

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



Definizione di grafo

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Un grafo è una coppia $G = (V, E)$ di insiemi tale che $E \subseteq [V]^2$; quindi, gli elementi di E sono sottoinsiemi a 2 elementi di V.
- Gli elementi di V sono i vertici (o nodi o punti) del grafo G, gli elementi di E sono i suoi archi.
- Ogni arco coinvolge due elementi che appartengono all'insieme V.

Insiemi dei vertici e degli archi

- L'insieme dei vertici di un grafo G è denotato come:

$$V(G) = \{v_1, v_2, \dots, v_n\}$$

- L'insieme degli archi di un grafo G è denotato come:

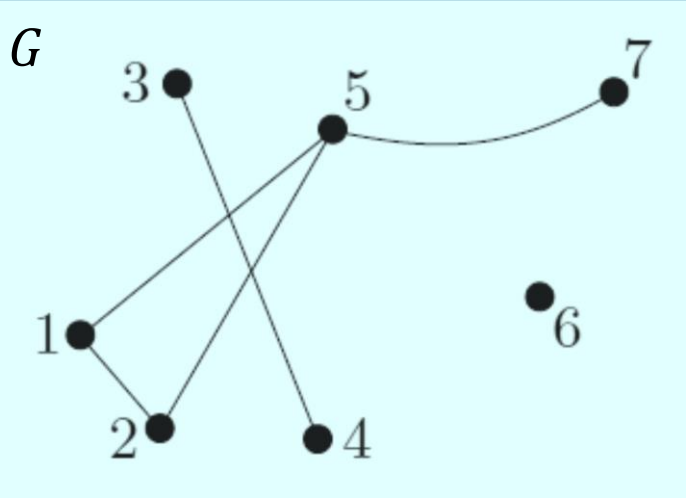
$$E(G) = \{e_1, e_2, \dots, e_m\}$$

- Denotiamo con $e = \{a, b\}$ (o $e = (a, b)$), un arco tra due vertici qualsiasi a e b.

Rappresentazione di un grafo

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Il modo usuale per rappresentare un grafo consiste nel disegnare un punto per ogni vertice e unire due di questi punti con una linea se i due vertici corrispondenti formano un arco.



Il grafo in figura ha come insieme dei vertici:

$$V(G) = \{1, 2, \ldots, 7\}$$

Ha come insieme degli archi:

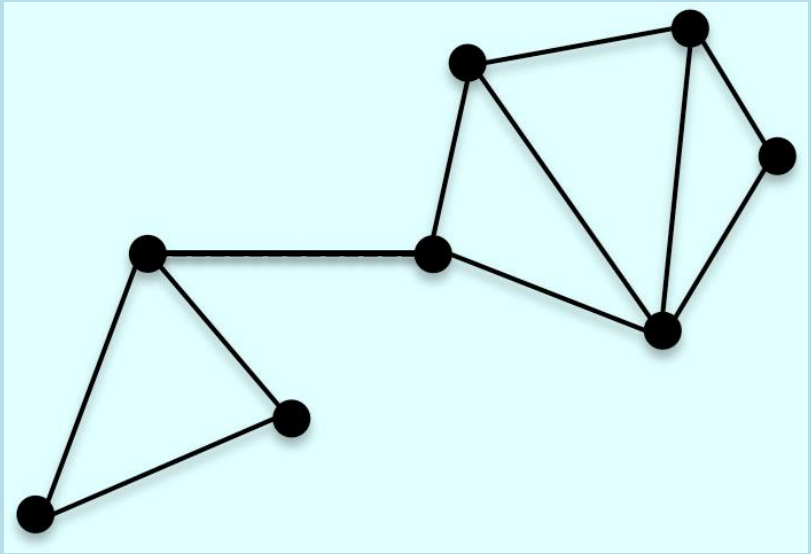
$$E(G) = \{\{1,2\},\{1,5\},\{2,5\},\{3,4\},\{5,7\}\}$$

O, con notazione alternativa:

$$E(G) = \{(1,2),(1,5),(2,5),(3,4),(5,7)\}$$

Grafo connesso

- In teoria dei grafi, un grafo $G = (V, E)$ è detto **connesso** se, per ogni coppia di vertici $(u, v) \in V$, esiste un **cammino** che collega u a v.



Ordine e dimensione di un grafo

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Il numero di vertici di un grafo rappresenta il suo **ordine**, denotato come:

$$|G| = |V(G)| = n.$$

- Il numero degli archi di un grafo rappresenta la sua **dimensione**, denotata come:

$$||G|| = |E(G)| = m.$$

Alcune statistiche

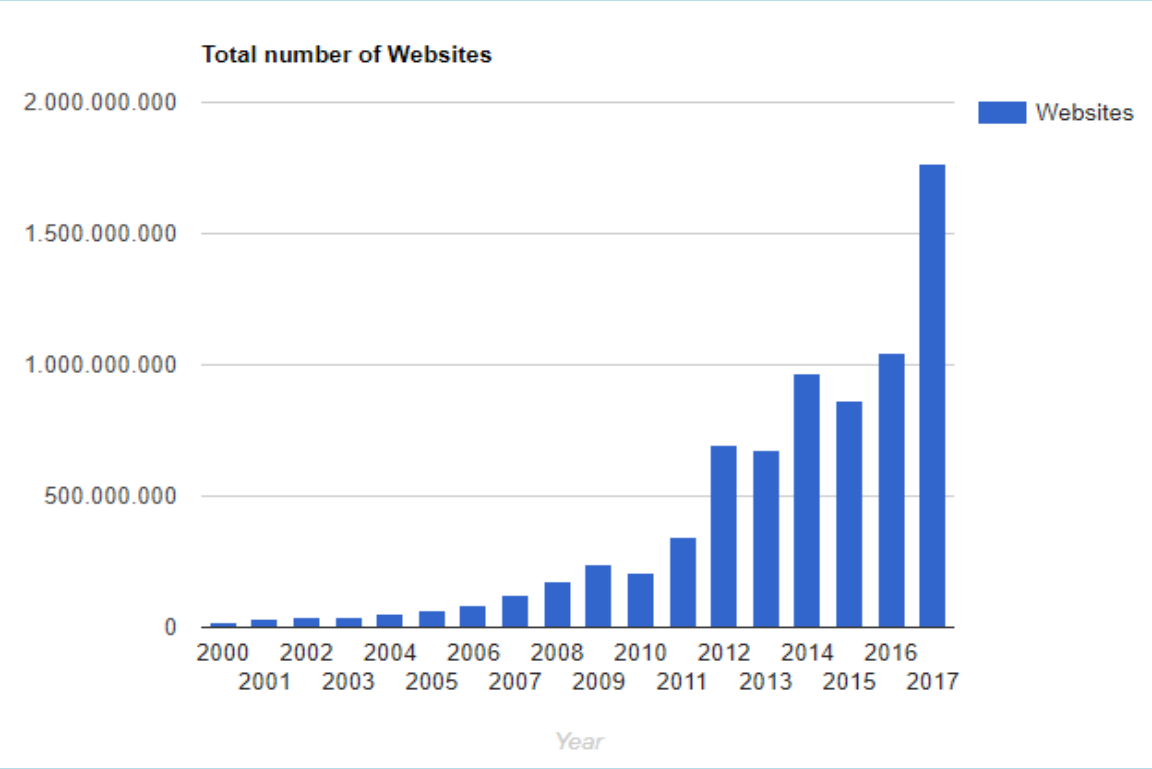
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- **Millioni di server** interconnessi (gestiti da istituzioni/compagnie diverse).
- A luglio 1999, erano stimati un numero di siti pari a 3 milioni.
- Crescita esponenziale.
- Ci sono circa 1.7 **miliardi di siti Web** oggigiorno.
  - Di questi, circa meno di 200 milioni sono attivi.
  - Il miliardo di siti fu raggiunto a settembre 2014.

http://www.internetlivestats.com/total-number-of-websites/

Numero di siti Web

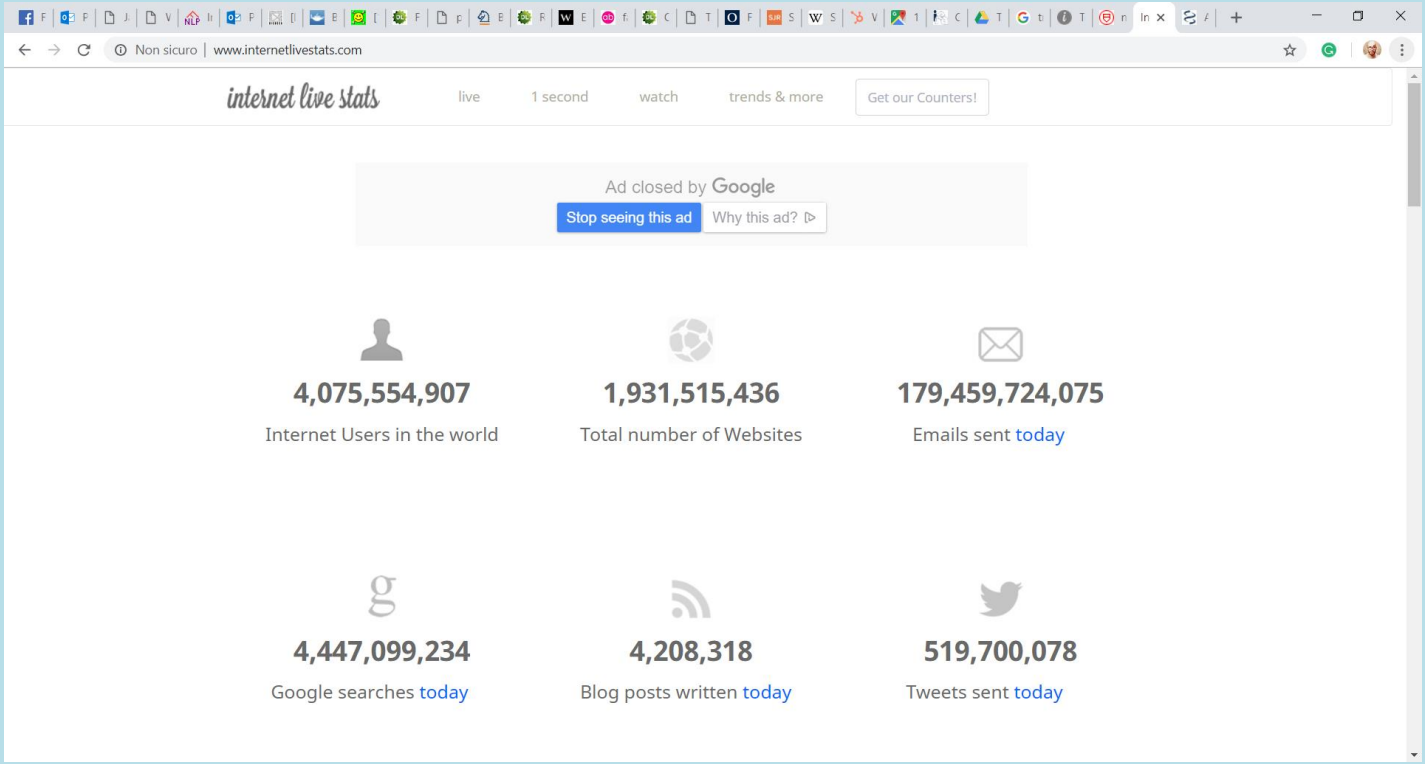
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



http://www.internetlivestats.com/total-number-of-websites/

Altre statistiche

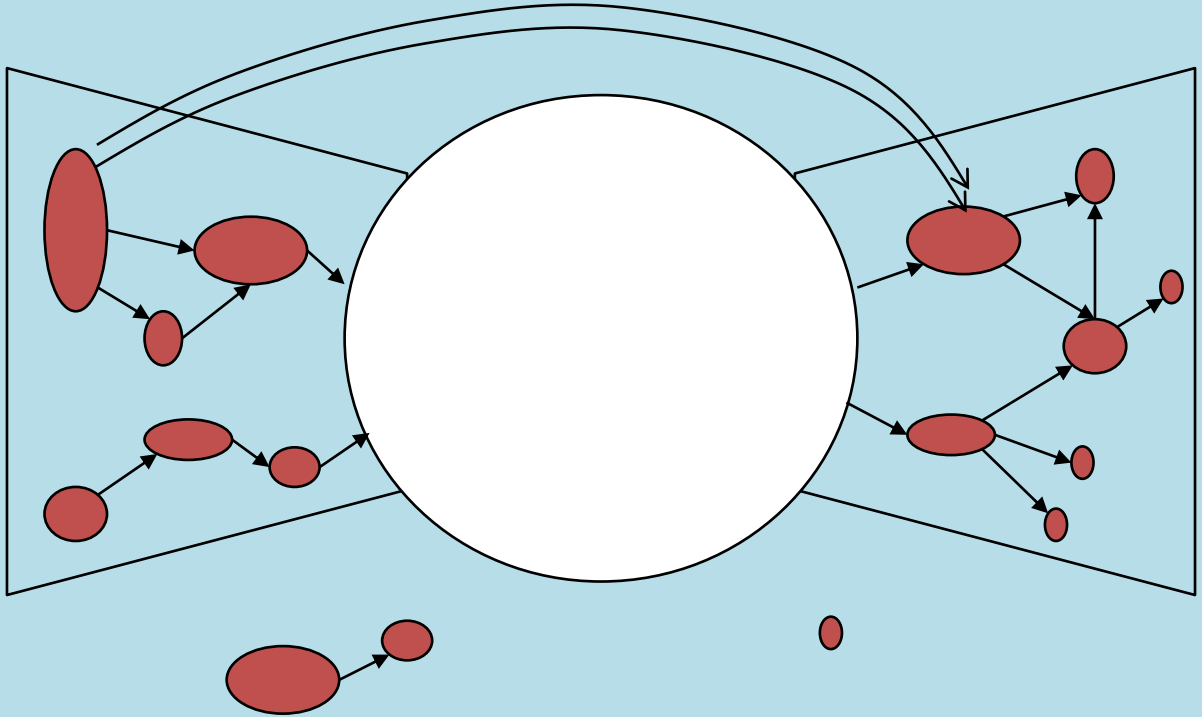
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



http://www.internetlivestats.com/

La struttura a «cravattino» del Web

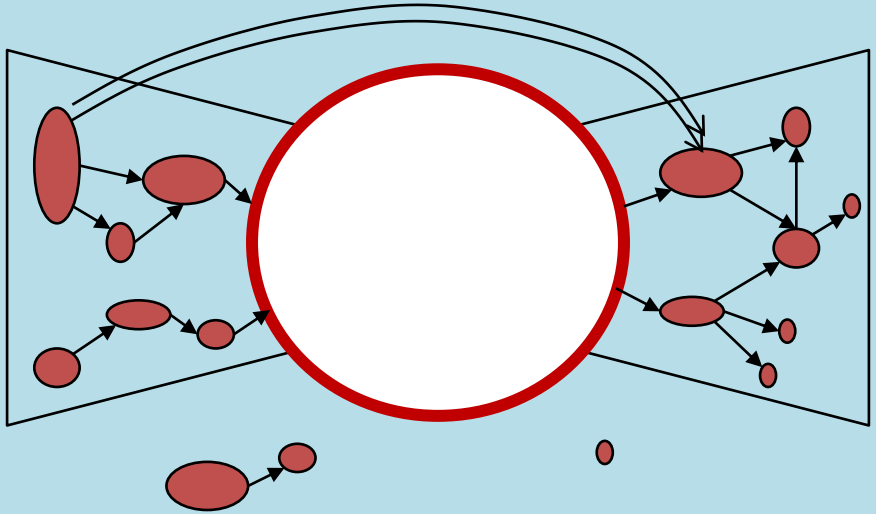
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE



La struttura a «cravattino» del Web
La componente gigante

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

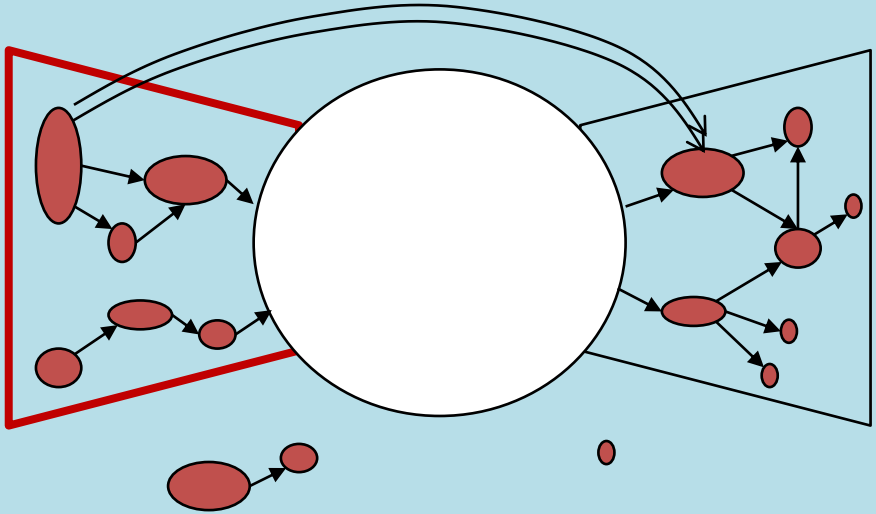
- Componente gigante:
  - Comprende la maggior parte delle pagine Web.



La struttura a «cravattino» del Web
Le componenti «sorgente»

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

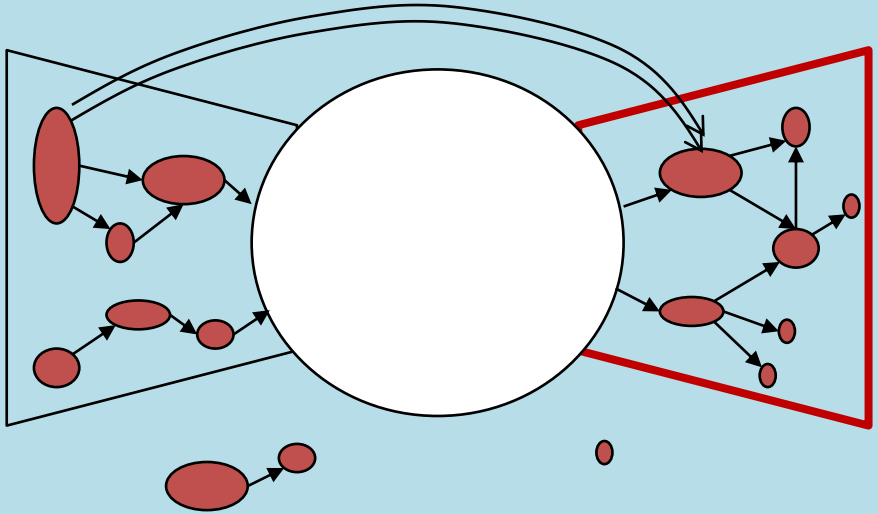
- Componenti «sorgente» (circa 1/5 delle pagine Web):
  - le pagine puntano direttamente o indirettamente verso la componente gigante MA …
  - … non sono raggiungibili dalla componente gigante
  - sono le pagine «reiette».



La struttura a «cravattino» del Web
Le componenti «pozzo»

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

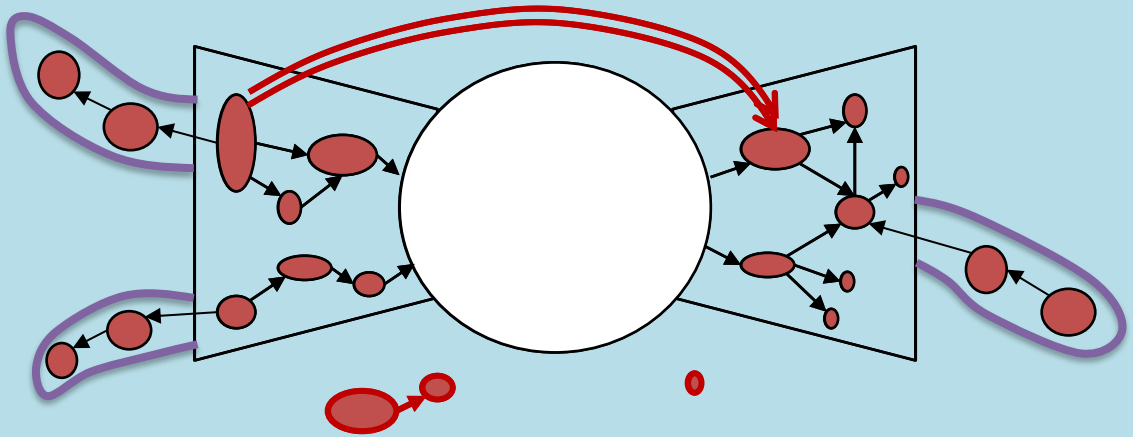
- Componenti «pozzo» (circa 1/5 delle pagine Web):
  - sono raggiungibili dalla componente gigante MA
  - da esse non si può tornare indietro;
  - in questa parte rientra la maggior parte dei documenti (senza link).



La struttura a «cravattino» del Web
Le componenti «isolate»

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Componenti «isolate», tentacoli e tubi:
  - non sono raggiungibili dalla componente gigante;
  - da esse non si raggiunge la componente gigante;
  - ci sono collegamenti fra sorgenti e pozzi che non passano per la componente gigante (tubi).



Accesso all'informazione sul Web

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Come è possibile **trovare le informazioni** all'interno del grafo del Web?
- **Lezione dedicata**:
  - Il problema del sovraccarico informativo.
  - Tutte le pagine sono «accessibili»? Il «dark Web» e il «deep Web».
  - I sistemi per il reperimento dell'informazione sul Web.
    - Motori di ricerca.
    - Sistemi di raccomandazione.

Copyright ®

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

- Queste slide fanno parte del corso “Strumenti e Applicazioni del Web”.

-

-