



# Elementi di Informatica

## Il World Wide Web: fondamenti

# Obiettivo della lezione

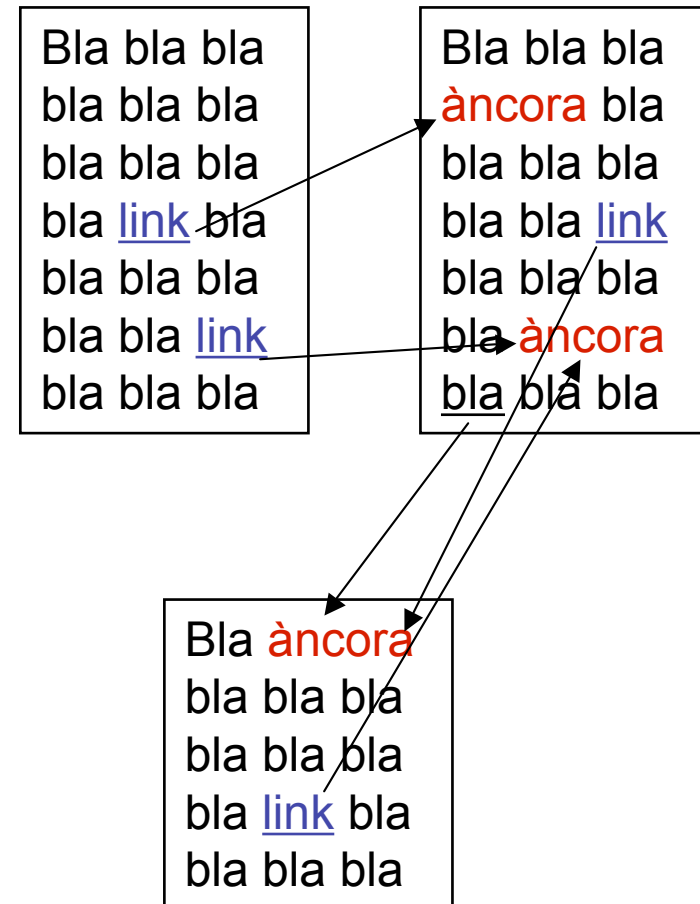
- La struttura del Web
- Il browser
- I linguaggi del Web: HTML e XML
- Servizi basati sul Web
  - Es. Blog, RSS, Wiki

# La storia

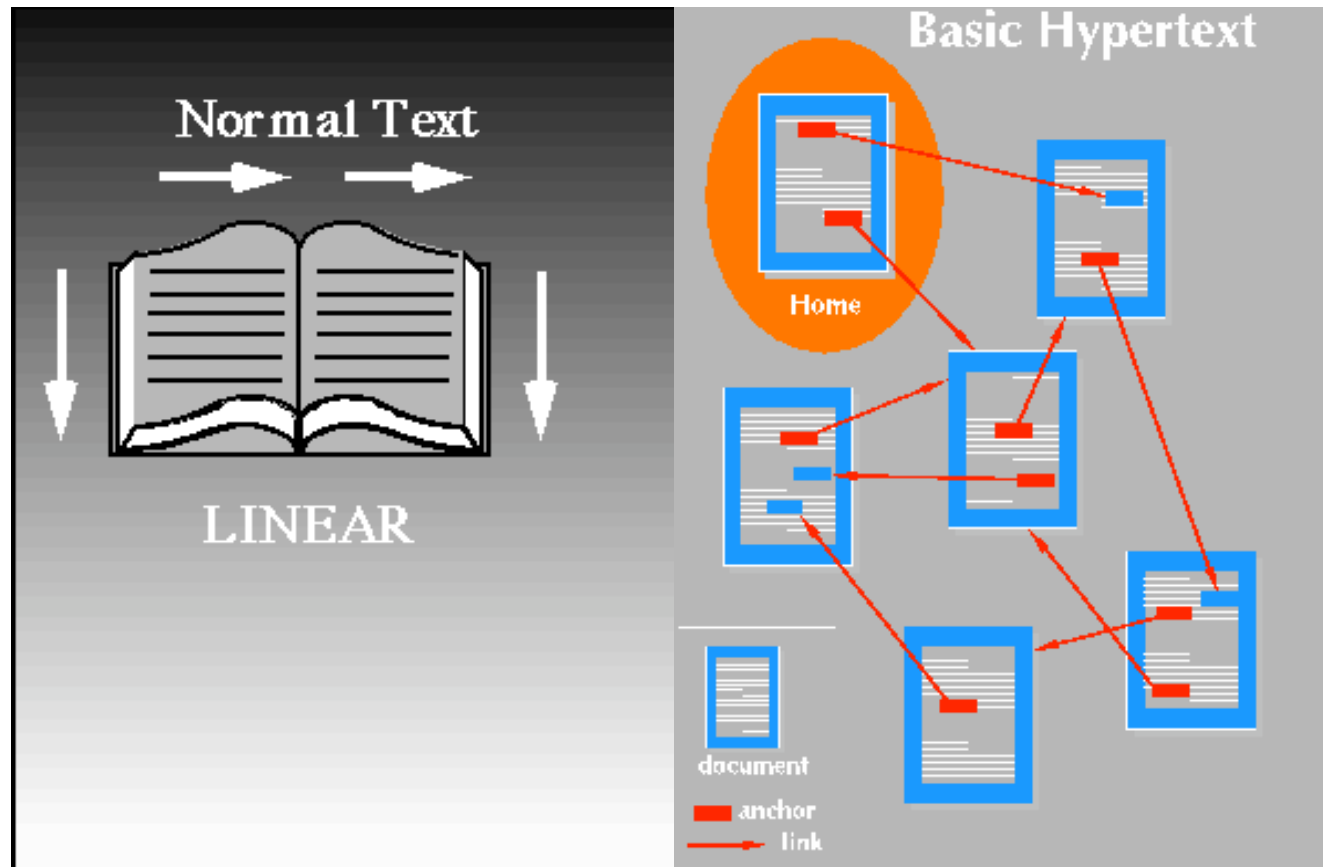
- 1989: T.Berners Lee al CERN progetta un sistema ipertestuale distribuito
- 1993: viene rilasciato Mosaic, il primo browser
- 1994: nasce il Consorzio WWW, che regola gli standard del Web
- 1995: Sun rilascia il linguaggio Java
- 1998: Nasce Google
- 2001: Nasce Wikipedia

# Ipertesti

- Un ipertesto è un testo che contiene riferimenti (link) navigabili a (punti interni ad) altri testi, in modo da permettere la lettura non lineare



# Iper testo



# Memex

“Un memex è un dispositivo in cui una persona memorizza tutti i suoi libri, dischi, e messaggi, e che è meccanizzato in modo che possa essere consultato con grande rapidità e flessibilità”

- Ricerca su tutto il testo, note audio e testuali, link

*Vannevar Bush, As We May Think, 1945*



# Memex di V. Bush

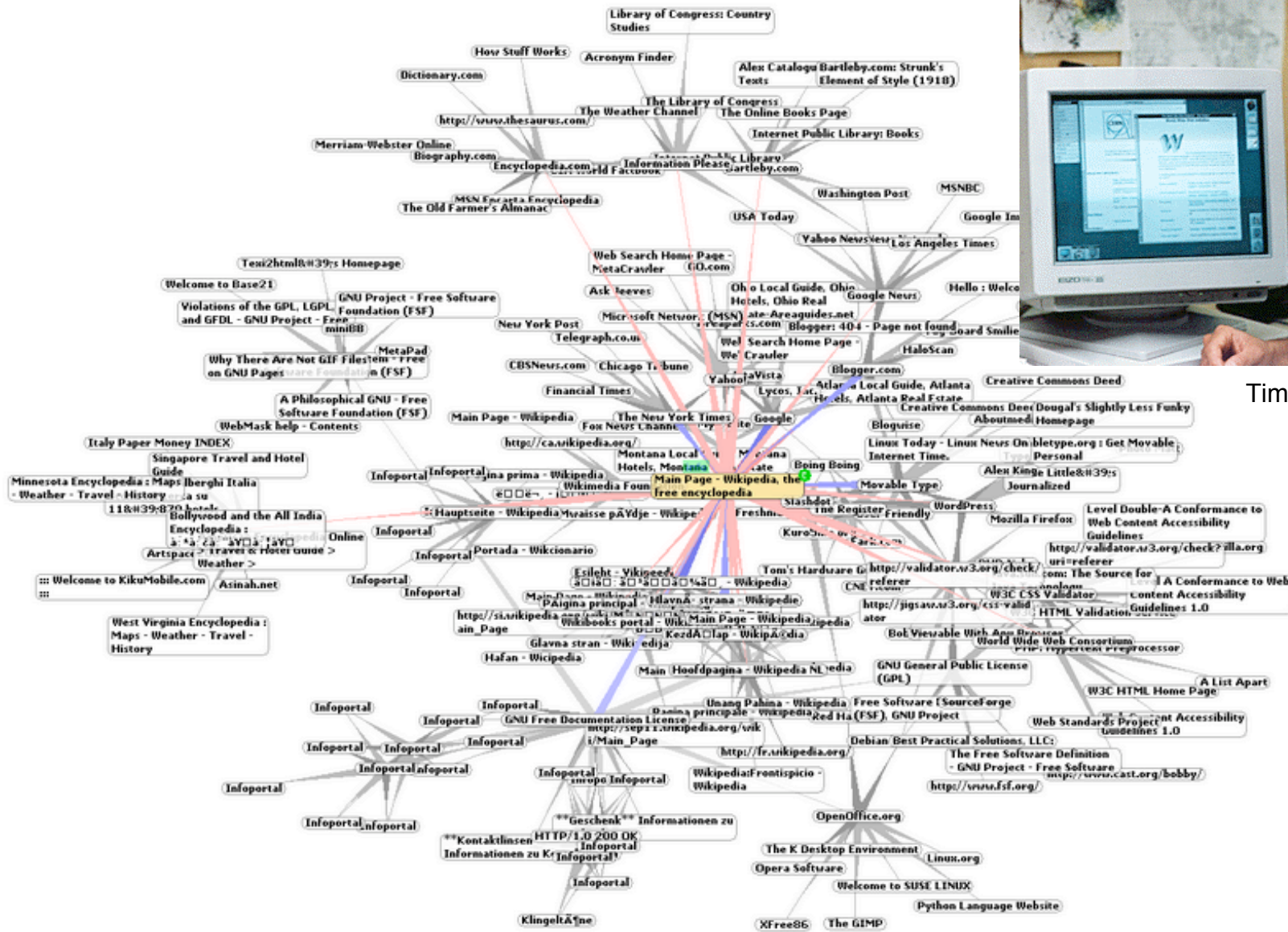
Consider a future device for individual use....



# World Wide Web



Tim Berners-Lee

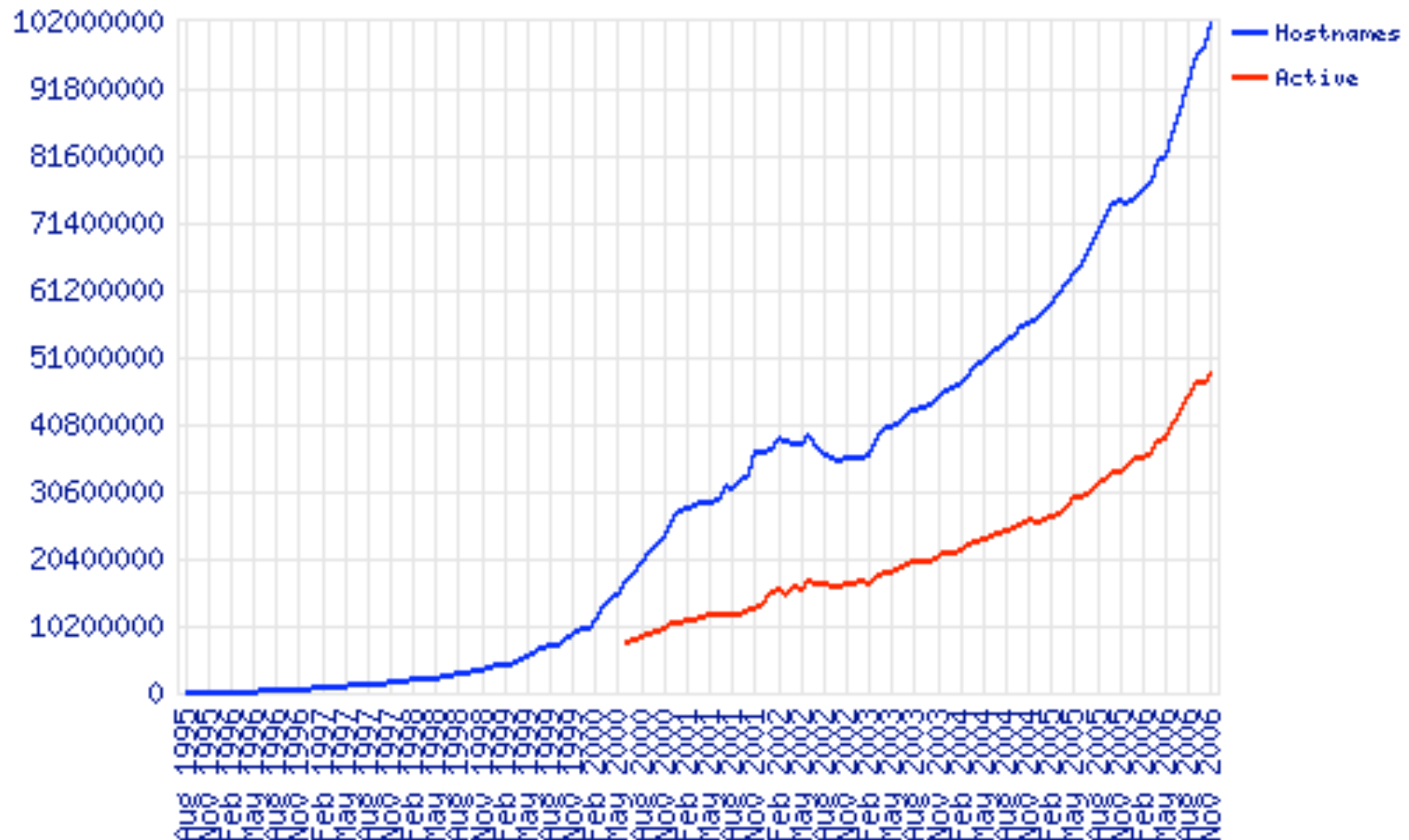




# Quanto è grande il Web?

- Oltre 100 milioni di siti
- Stima: il Web contiene oltre 100 miliardi di pagine
- Stima: cresce al ritmo di 10 milioni di pagine al giorno
- Stima: almeno 100 milioni di pagine vengono modificate ogni giorno

# Nubero siti web (fonte [www.netcraft.com](http://www.netcraft.com))



# Cosa contiene il Web?

- Documenti (es. HTML, XML, JPG, .xls, ecc.)
- Documenti multimediali
- Documenti dinamici (es. Flash)
- Banche di dati
- Programmi
- Servizi (es. Wiki, feed RSS, ecc.)

Chiamiamo tutto questo "risorse Web"

# Una pagina HTML

```
<HTML><HEAD><TITLE>Paolo Ciancarini</TITLE></HEAD>
<BODY><BODY BGCOLOR="#FF
```

```
<TABLE>
```

```
<TR><TD><H4>
```

```
<A HREF="../index.html">Affiliazi
<A HREF="wwwpages/teaching.h
<A HREF="wwwpages/research.f
<A HREF="wwwpages/seminari/li
<A HREF="wwwpages/personal.h
<A HREF="ftp://ftp.cs.unibo.it/pub
<A HREF="wwwpages/curr.pdf">|
<BR></H4></TD>
```

[Affiliazione](#)  
[I miei insegnamenti](#)  
[Interessi di ricerca](#)  
[Alcuni miei seminari](#)  
[Interessi personali](#)  
[I miei articoli e rapporti](#)  
[Il mio curriculum](#)



[Affiliation](#)  
[My teaching duties](#)  
[Research Interests](#)  
[A few talks of mine](#)  
[Personal Interests](#)  
[Technical Reports](#)  
[My CV \(in Italian\)](#)

```
<TD><IMG ALIGN=LEFT HSPACE
<TD><H4>
```

```
<A HREF="../index.html">Affiliatic
<A HREF="wwwpages/teaching.h
<A HREF="wwwpages/research.f
<A HREF="wwwpages/seminari/li
<A HREF="wwwpages/personal.h
<A HREF="ftp://ftp.cs.unibo.it/pub
<A HREF="wwwpages/curr.pdf">|
```

## Find me

```
<br></H4></TD></TR></TABLE><
```

prof. Paolo Ciancarini Dipartimento di Scienze dell'Informazione - Univ. di Bologna  
Mura Anteo Zamboni 7, 47127 Bologna - Italy

```
<hr><H2>Find me</H2>
```

```
<HR>prof. Paolo Ciancarini
Dipartimento di Scienze dell'Inform
Mura Anteo Zamboni 7, 47127 Bol
Tel. +39051 2094506 <br>
Fax: +39051 2094510 <br>
<HR>Last Updated: May 5, 2005
```

Tel. +39051 2094506  
Fax: +39051 2094510

Last Updated: May 5, 2005

```
<HR>
```

```
</BODY></HTML>
```

# Publicare una pagina

- Occorre preparare il file della pagina, per esempio con un editor
- Occorre avere la disponibilità di uno spazio sul disco di uno host che contiene un server Web
- Occorre inserire (*upload*) la pagina nello spazio del server

- Home
- Ateneo
- Studenti
- Offerta formativa
- Ricerca
- Relazioni Internazionali
- Biblioteche e musei
- Personale

### SERVIZI ONLINE

- Alma digital library
- Almalaurea
- AlmaWelcome!
- Autocertificazione Online
- Convenzioni e tirocini
- Erasmus online
- Exchange students
- Iscrizione agli esami
- La mia e-mail
- Liste di distribuzione
- Merchandising
- Newsletter studenti
- Piani di studio online
- Servizi bibliotecari
- Stradario di Bologna
- Università nel mondo
- Il mio Portale
- Per il Personale

Guida ai servizi

### STRUTTURE

- Aree amministrative
- Biblioteche
- Centri linguistici
- Centri interdipartimentali
- Centro di Bertinoro
- Collegio di Cina
- Collegio Superiore
- Comitato Pari Opportunità
- Centro Universitario Sportivo
- Dipartimenti
- Facoltà
- Fondazione Alma Mater
- Istituto di Studi Avanzati
- Poli della Romagna
- Residenza di Studi Superiori
- Scuole di specializzazione
- Scuole Superiori
- Sede di Buenos Aires
- Segreterie studenti
- Strutture di servizio
- URP - Relazioni con il pubblico

- [web Unibo](#)
- [Rubrica](#)
- [Altro](#)

in questo sito  in tutti i siti

Facoltà | Dipartimenti | Relazioni con il pubblico | Amministrazione generale

### IN PRIMO PIANO

#### [Piani di studio online](#)



Se devi presentare, aggiornare o modificare il tuo piano di studio è questo il periodo: consulta il sito della tua Facoltà oppure contatta l'Ufficio Didattico per avere maggiori dettagli riguardo scadenze e modalità di presentazione.



#### [Unibosearch al tuo servizio](#)

Scopri le risorse del motore di ricerca! Se cerchi qualcosa di specifico non fermarti alla ricerca semplice, ma sfrutta i filtri della ricerca avanzata.



#### [Puoi ancora immatricolarti!](#)

Fino al 28 dicembre 2006 sarà possibile immatricolarsi a qualunque corso a libero accesso. Leggi come fare.



#### [Trasferimento da un altro Ateneo](#)

Se hai fatto domanda di trasferimento da un altro Ateneo all'Università di Bologna, ricorda che hai tempo fino al 28 dicembre per pagare le tasse di trasferimento.



#### [Ti sei appena laureato?](#)

L'Unibo ti è vicina anche dopo la laurea. Consulta la vasta offerta di formazione e specializzazione per scegliere la strada che fa per te!

### MULTICAMPUS



#### [INFOsegreteria@online](mailto:INFOsegreteria@online)

Se non vuoi fare la fila in segreteria e vuoi trovare le risposte che cerchi via web, prova il servizio Supporto

## Servizi del portale



## Opportunità nella ricerca

Sono tante le opportunità offerte dall'Ateneo di Bologna ai giovani laureati interessati a proseguire il proprio percorso nella ricerca.

[Continua](#)



### Notizie



1 dicembre 2006

[La voce regina: diventa pubblico l'archivio di poesia sonora di Enzo Marinetti](#)

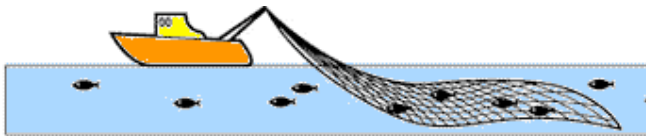
1 dicembre 2006

[Accordo di cooperazione culturale tra Libera e Unibo](#)

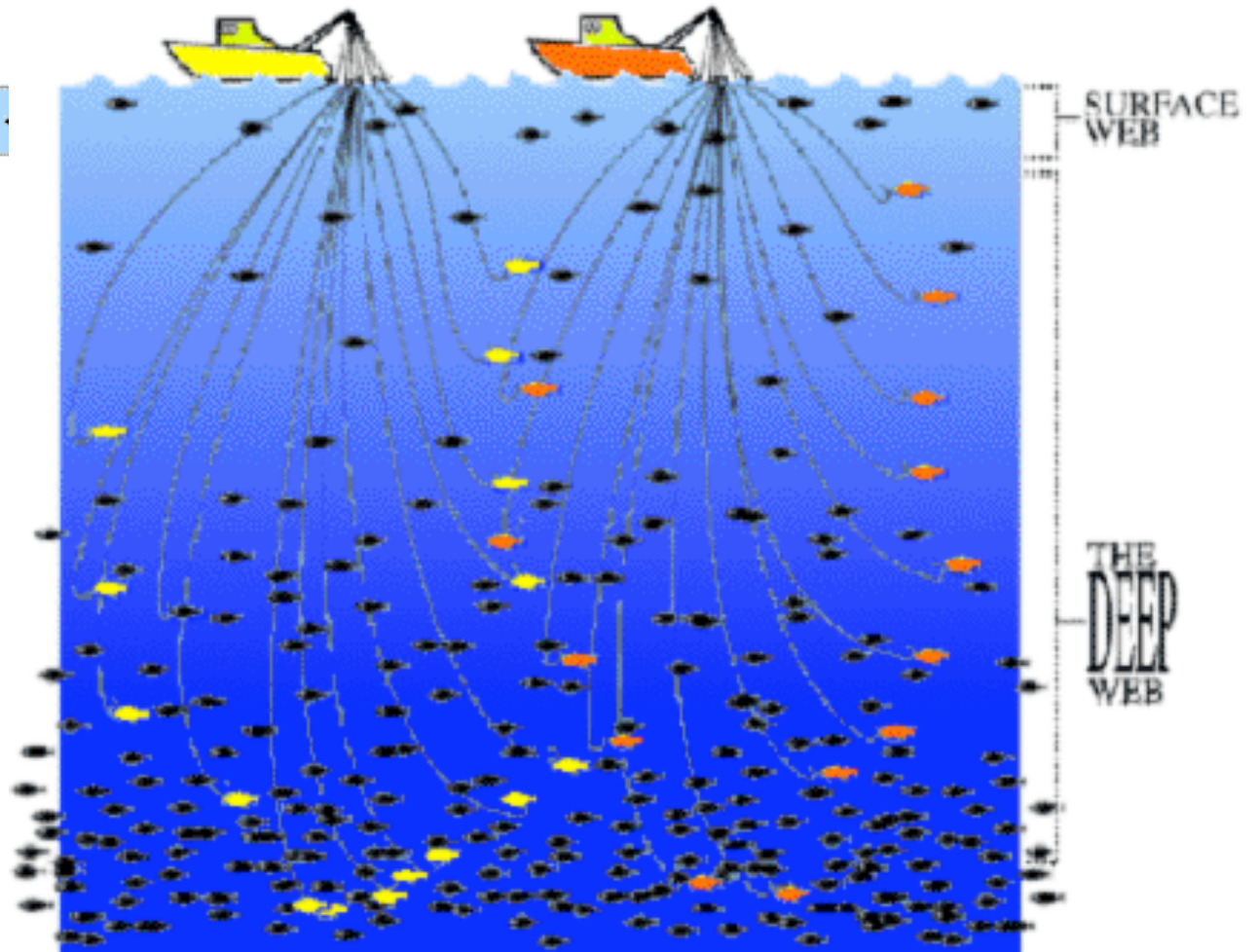
### Eventi



# Il "Web profondo"



Non tutti i documenti disponibili in rete sono collegati all'ipertesto globale, quindi non si possono trovare né "navigando" né coi motori di ricerca



# Come si trovano le risorse su Web

- Occorre conoscerne il nome (URL)  
oppure
- Occorre saper usare un motore di  
ricerca



# URL

- URL: Uniform Resource Locator
- Denota una risorsa accessibile via Web, di solito un file
- **Esempio:**

`http://www.cs.unibo.it/rossi/index.html`

protocollo

server

cammino

# HTTP

- HTTP sta per Hypertext Transfer Protocol
- E' un linguaggio di comandi che si scambiano i browser e i server del Web
- Esempio: "GET url" è il comando che un browser dà ad un server per dire "dammi la pagina di nome *url*"

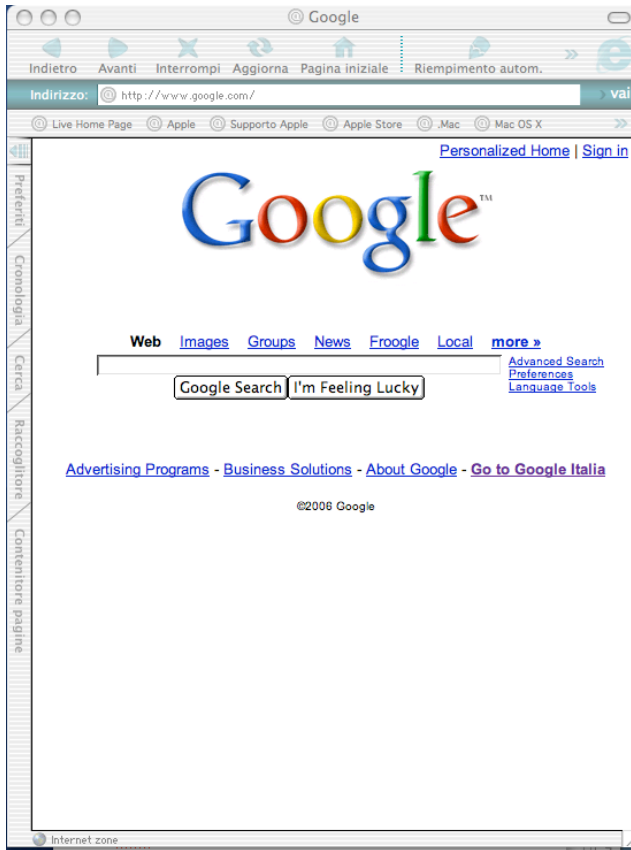
# Server Web

- Un server Web è un programma che gira su uno host connesso a Internet e risponde a comandi nel protocollo HTTP
- Il programma controlla uno spazio di pagine, detto "sito Web"
- Un famoso programma server Web open source è Apache

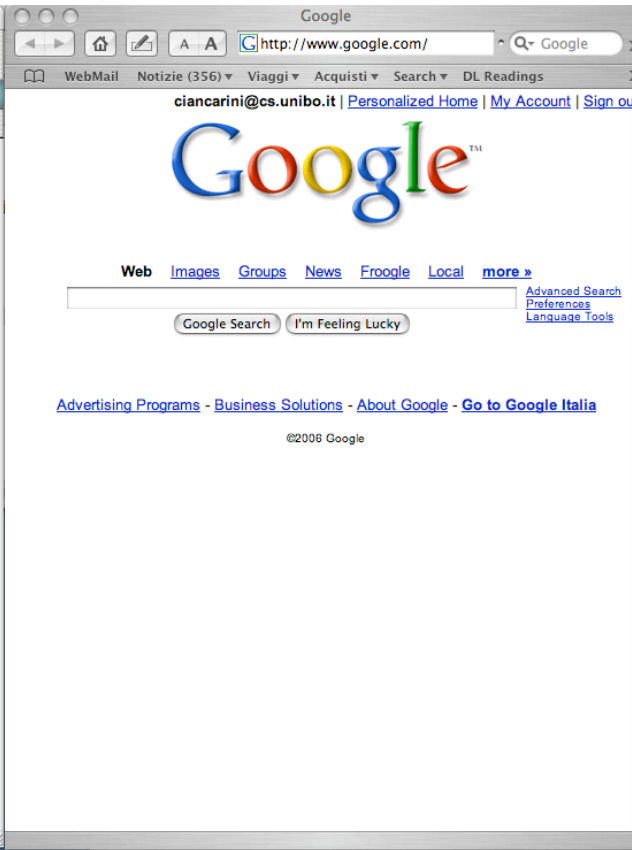
# Browser

- Un **browser** è un programma visualizzatore di pagine HTML, capace di parlare con server Web mediante il protocollo HTTP
- In sostanza permette di navigare ipertesti
- Ha molte altre funzioni, grazie ai plug-in (programmi aggiuntivi) e agli applet (programmi Java scaricabili come pagine Web)
- Ne esistono molti: Internet Explorer, Firefox, Opera, Safari, Mozilla, ecc.

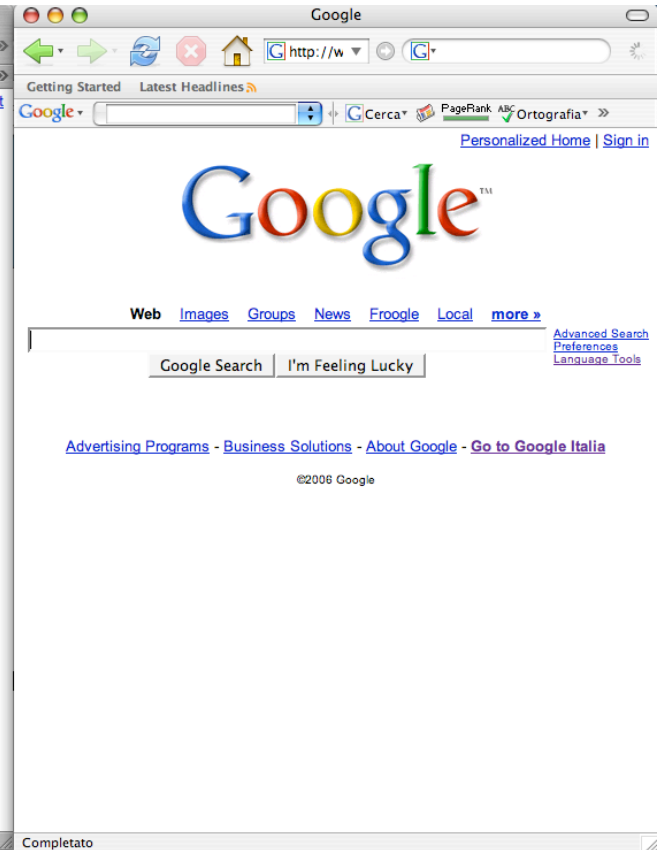
# Browser



Internet Explorer



Safari (Mac)



Firefox

# Configurare il browser

- Il browser dev'essere configurato
  - Aspetti di sicurezza (es. Java, cookies)
  - Home page su cui aprirsi
  - Aspetto dei caratteri
  - Dove registrare le pagine
  - Bookmarks
  - Configurazione di lingua
  - Riempimento automatico
  - Ogni browser sa visualizzare direttamente file html e ascii: occorre definire come aprire i documenti in altri formati, aggiungendo plug in specifici

# Sicurezza nel browser

Contenuti Web:  Abilita plug-in

Abilita Java

Abilita JavaScript

Blocca le finestre a comparsa

Accetta Cookie:  Sempre

Mai

Solo dai siti che stai consultando

Per esempio, non dagli sponsor presenti in quei siti

Mostra Cookie

Chiedi prima di inviare un modulo non sicuro ad un sito sicuro

Abilita i Controlli censura

# Navigazione privata



## Sei sicuro di voler attivare la navigazione privata?

Se la navigazione privata è attivata, le pagine web non vengono aggiunte alla cronologia, gli elementi vengono rimossi automaticamente dalla finestra dei download, le informazioni per il riempimento automatico non vengono registrate (neanche i nomi e le password), e le ricerche non vengono aggiunte al menu a comparsa della casella di ricerca Google. Finché non chiudi la finestra, puoi ancora fare clic sui pulsanti per andare indietro o avanti e tornare alle pagine Web che hai aperto.



Annulla

OK



# Il Web

- Il Web è fatto di **server** connessi a Internet, che gestiscono insiemi di documenti o banche dati identificati da URL
- I documenti e le banche dati hanno un contenuto che viene visualizzato dal browser
- Se il contenuto è codificato in HTML può riferire altri documenti mediante hyperlinks
- L'accesso ai documenti è *diretto*, se se ne conosce l'URL, o *mediato* da un motore di ricerca, che viene interrogato per “trovare” il documento

# Navigare il Web

- Un browser su una macchina "cliente" mostra un documento all'utente
- L'utente clicca su un link
- Il link viene trasformato in indirizzo IP
- La macchina cliente contatta il server all'indirizzo IP
- Il server risponde al cliente dando la nuova pagina (se la pagina non esiste dà errore)
- Il browser mostra all'utente la nuova pagina

Errore (pagina non esiste)



## Not Found

The requested URL `/prova` was not found on this server.

---

# HTML

- HTML sta per HyperText Markup Language
- Un linguaggio di markup è un linguaggio che descrive le caratteristiche grafiche e strutturali di un documento

# Schema base di un documento HTML

```
<HTML>
```

```
<HEAD>
```

```
  <TITLE>Esempio</TITLE>
```

```
</HEAD>
```

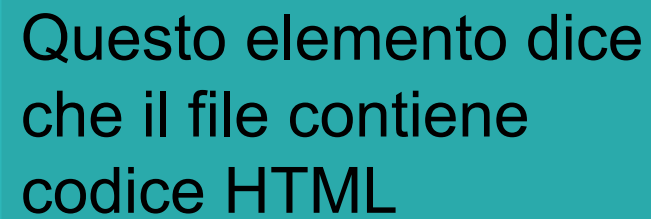
```
<BODY>
```

```
</BODY>
```

```
</HTML>
```

# Schema base

<HTML>



Questo elemento dice  
che il file contiene  
codice HTML

</HTML>

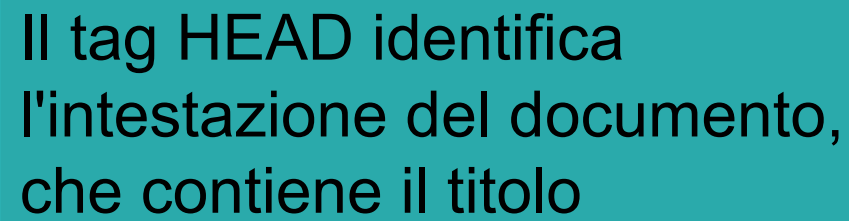
# Schema base

<HTML>

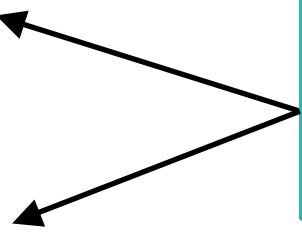
<HEAD>

</HEAD>

</HTML>

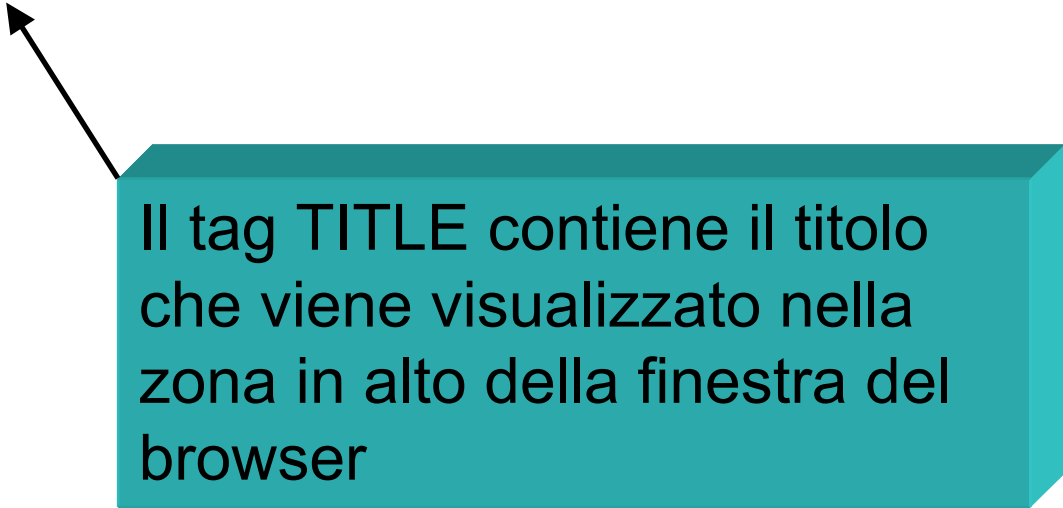


Il tag HEAD identifica  
l'intestazione del documento,  
che contiene il titolo



# Schema base

```
<HTML>  
<HEAD>  
  <TITLE>Esempio</TITLE>  
</HEAD>  
  
</HTML>
```



Il tag TITLE contiene il titolo che viene visualizzato nella zona in alto della finestra del browser



# Schema base

<HTML>

<HEAD>

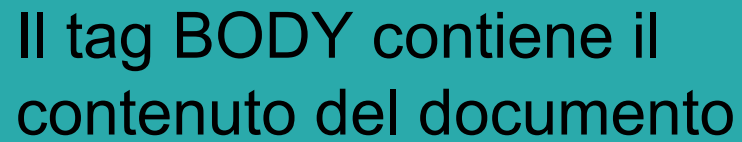
<TITLE>Esempio</TITLE>

</HEAD>

<BODY>

</BODY>

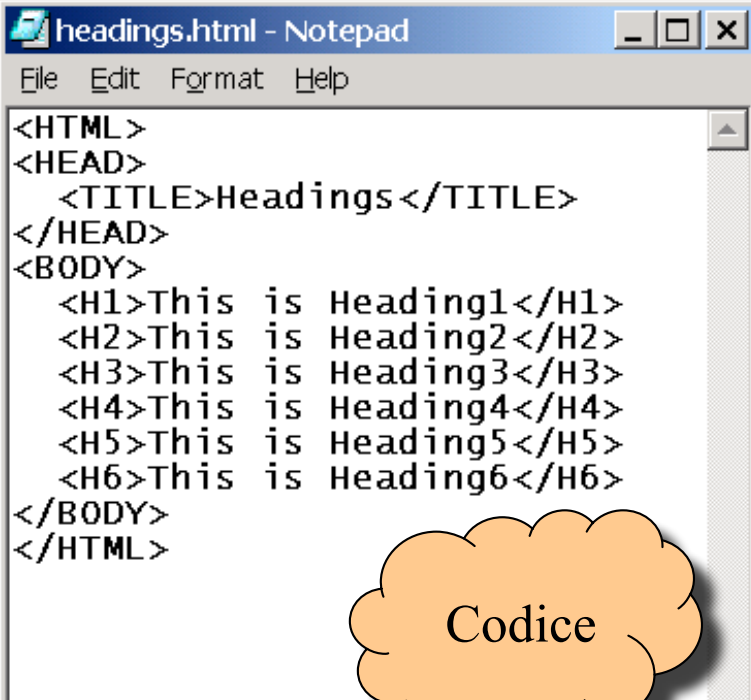
</HTML>



Il tag BODY contiene il  
contenuto del documento

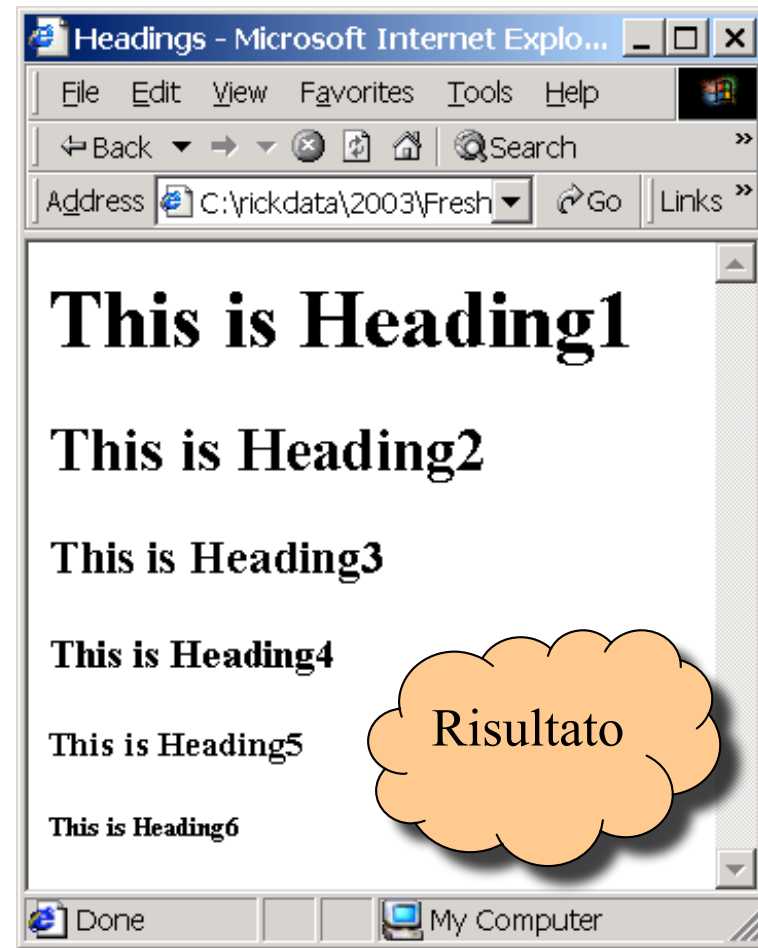
# Titoletti

- HTML ha sei livelli di titoletti
- Il primo livello è <H1>



```
<HTML>
<HEAD>
  <TITLE>Headings</TITLE>
</HEAD>
<BODY>
  <H1>This is Heading1</H1>
  <H2>This is Heading2</H2>
  <H3>This is Heading3</H3>
  <H4>This is Heading4</H4>
  <H5>This is Heading5</H5>
  <H6>This is Heading6</H6>
</BODY>
</HTML>
```

Codice



# Controllare l'aspetto grafico

<P> Paragraph Tag </P>

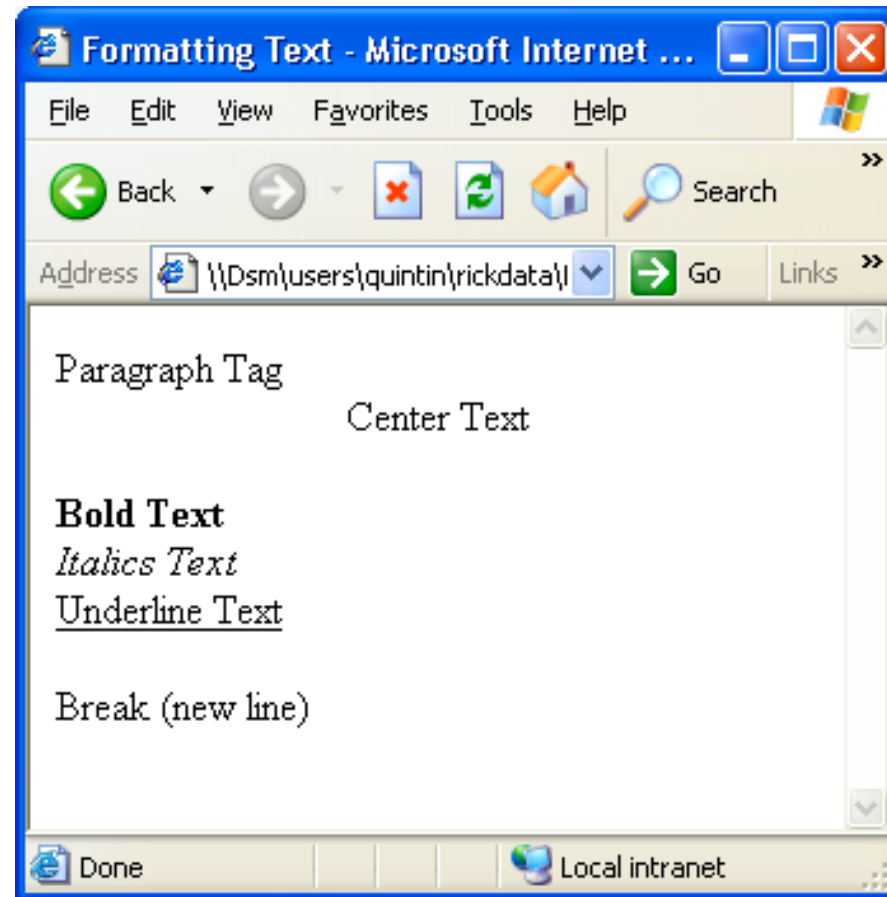
<CENTER> Center  
Text</CENTER>

<B>Bold Text</B>

<I>Italics Text</I>

<U>Underline Text</U>

<BR>Break (new line)



# Cos'è un *Markup Language*

- Un linguaggio capace di descrivere singoli aspetti di un documento digitale
- Specifica informazioni di struttura e/o layout
- Collega informazioni entro il documento
- Collega insieme diversi documenti
- Può incorporare componenti multimediali
- Può esprimere informazioni utili per l'indicizzazione del documento

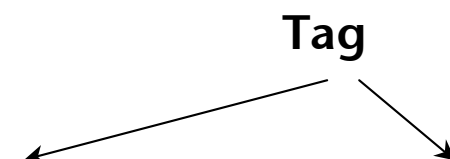
# Markup

- Manoscritto (a macchina)

Questa è una nozione importante.

- HTML

Questa è una nozione `<em>importante</em>`.

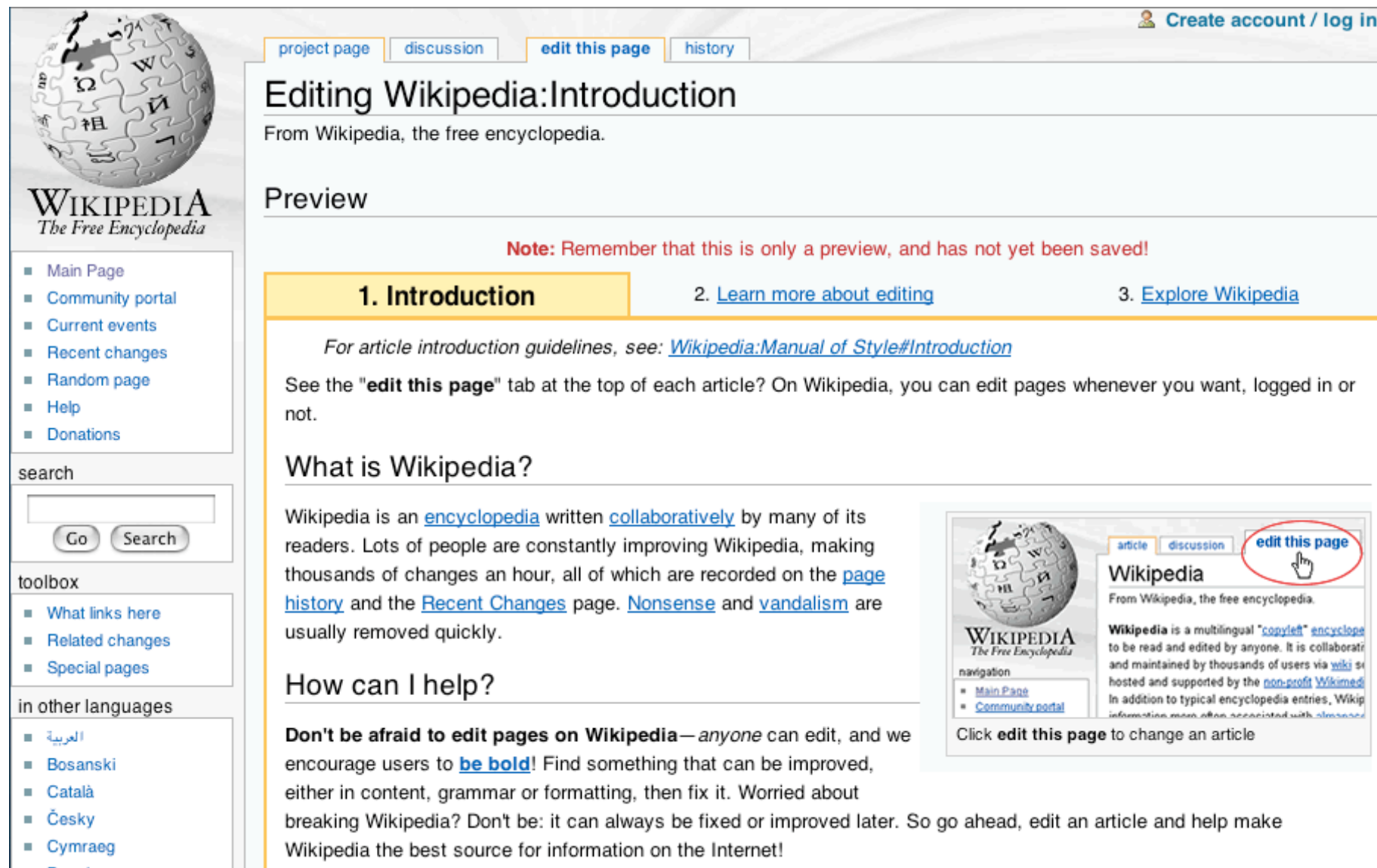


```
graph TD; Tag --> em["<em>"]; Tag --> em_slash["</em>"];
```

- XML

```
<block>
  Questa è una nozione
  <emp>importante</emp>.
</block>
```

# Wiki: la minimizzazione del markup presentazionale



[project page](#) [discussion](#) [edit this page](#) [history](#) 👤 Create account / log in

## Editing Wikipedia: Introduction

From Wikipedia, the free encyclopedia.

### Preview

**Note:** Remember that this is only a preview, and has not yet been saved!

- 1. Introduction**
- [2. Learn more about editing](#)
- [3. Explore Wikipedia](#)

*For article introduction guidelines, see: [Wikipedia:Manual of Style#Introduction](#)*


See the "**edit this page**" tab at the top of each article? On Wikipedia, you can edit pages whenever you want, logged in or not.

### What is Wikipedia?

Wikipedia is an [encyclopedia](#) written [collaboratively](#) by many of its readers. Lots of people are constantly improving Wikipedia, making thousands of changes an hour, all of which are recorded on the [page history](#) and the [Recent Changes](#) page. [Nonsense](#) and [vandalism](#) are usually removed quickly.

### How can I help?

**Don't be afraid to edit pages on Wikipedia**—*anyone* can edit, and we encourage users to **be bold**! Find something that can be improved, either in content, grammar or formatting, then fix it. Worried about breaking Wikipedia? Don't be: it can always be fixed or improved later. So go ahead, edit an article and help make Wikipedia the best source for information on the Internet!



article [discussion](#) [edit this page](#)

### Wikipedia

From Wikipedia, the free encyclopedia.

**Wikipedia** is a multilingual "copyleft" [encyclopedia](#) to be read and edited by anyone. It is collaborati and maintained by thousands of users via [wiki](#) s; hosted and supported by the [non-profit Wikimedia](#). In addition to typical encyclopedia entries, [Wikip information](#) *more often associated with classes*

Click **edit this page** to change an article

# RSS

- Linguaggio per pagine Web "machine-readable"
- Formato XML per titoli e descrizioni di notizie giornalistiche commerciabili (syndicate)
- Sviluppato da Netscape, adottato dai siti blog
- Usato da:
  - Utenti con browser specifici
  - Webmaster che includono nelle loro pagine dinamiche titoli da altri siti
  - Aggregatori di notizie

# Esempio RSS

The screenshot shows a web browser window with the address bar containing the RSS feed URL: `feed://www.corriere.it/rss/homepage.xml`. The browser's address bar also shows the RSS icon and the Google search engine. The page title is "Corriere.it - Cronache" and the number of items is "43 in tot". The main content area displays two news items: "Abusava della nipotina, arrestato 60enne" (dated "Oggi, 08:40 p.") and "Pamuk, la Turchia rinuncia al processo" (dated "Oggi, 06:28 p."). On the right side, there is a search box labeled "Cerca negli articoli:" and a field for "Lunghezza articolo:".

The screenshot shows the Yahoo! Mio Yahoo! Mail interface. The top navigation bar includes "Yahoo! Mio Yahoo! Mail" and the "MIO YAHOO! Entra" logo with links for "Nuovo utente?" and "Registrati". The main content area features a section titled "Aggiungi Corriere.it - Cronache a Mio Yahoo!" with a yellow "Aggiungi" button. Below this, a text block states: "Con Corriere.it - Cronache e tutti gli altri feed disponibili, la tua pagina Mio Yahoo! è sempre aggiornata. Cos'è Mio Yahoo!? [Altre info.](#)". A list of news items from the "Corriere.it - Cronache" feed is displayed:

- [Incontro al Quirinale Ciampi-Berlusconi](#) - 8 hours ago
- [Alitalia: annullato lo sciopero di lunedì](#) - 10 hours ago
- [Georgia: gas e linee elettriche interrotte](#) - 10 hours ago
- [Aviaria: caso sospetto in Francia](#) - 7 hours ago
- [Tragedia sulle Apuane, morti tre alpinisti](#) - 3 hours ago

**Aiutare i gorilla del Congo riciclando cellui...** A. Mu. Ieri, 03:18 p.

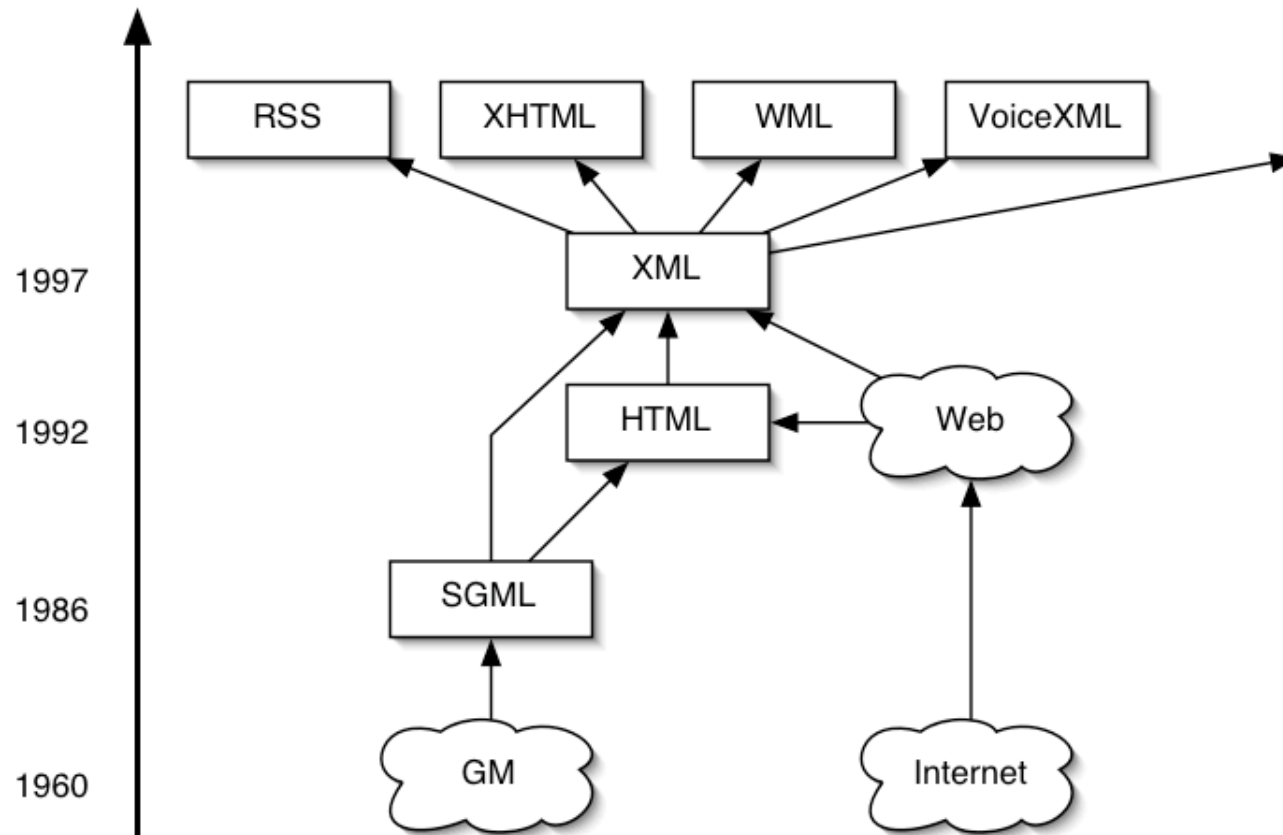
**Galliani: «Vogliamo giustizia»** Ieri, 02:55 p.

**Fassino: Berlusconi ha fallito, cerca la rissa** Ieri, 02:54 p.

**«Detoscanizzare l'Italia? Solo una battuta»** Ieri, 02:53 p.



# Storia dei Linguaggi di Markup



# Esempi di linguaggi di markup

- RUNOFF, troff, nroff
- SGML Standard Generalized Markup Language
- HTML HyperText Markup Language
- GML Geography Markup Language
- TeX, LaTeX
- LMNL Layered Markup Annotation Language
- RTF Rich Text Format
- WML Wireless Markup Language
- VRML Virtual Reality Markup Language
- XML eXtensible Markup Language

# Versioni di HTML

- Quando nel 1989, al CERN, venne inventato il WWW, si usava un browser **non-grafico** che gestiva una versione detta HTML 0
- HTML 1 (1993) fu quello supportato dal browser Mosaic; la principale innovazione furono la possibilità di inserire **immagini** e la gestione grafica della navigazione ipertestuale
- HTML 2 (1994) venne definito usando SGML; introdusse una gestione semplice delle **tabelle** e degli **script**
- HTML 3.0/3.2 (1996) aumentò le funzionalità ipertestuali; le maggiori novità furono le **tabelle**, i **font colorati**, e soprattutto gli **applet Java**
- HTML 4 (1998) è la versione più recente e attuale dello standard; introdusse i **fogli di stile**

# Cascading StyleSheets (CSS)

- Un foglio di stile (o **stylesheet**) è un insieme di regole che istruiscono un browser su come presentare (a video, a stampa) un documento HTML
- Ci sono vari modi di collegare regole di stile a documenti HTML
- Un modo semplice consiste nell'uso dell'elemento **STYLE** di HTML, che viene messo nell'intestazione (**HEAD**) del documento e contiene le regole di stile della pagina
- Ogni regola si compone di due parti: *selector { property: value }*
  - un selettore (*selector*): di solito un elemento HTML come **BODY**, **P**, o **EM**;
  - lo stile da applicare al selettore, definito da coppie *property/value*
- Ogni elemento può avere parecchie proprietà (*properties*)
- Ogni proprietà ha un valore che descrive come va presentato il selettore

# Esempio: stylesheet CSS

- Definiamo le proprietà *color* e *font-size* per gli elementi H1 e H2:

```
<HEAD>
<TITLE>CSS Example</TITLE>
<STYLE TYPE="text/css">
  H1 { font-size: x-large; color: red }
  H2 { font-size: large; color: blue }
</STYLE>
</HEAD>
```

- Questo stylesheet dice al browser di mostrare i titoli di livello 1 con fonte rossa *extra-large*, mentre per i titoli di livello 2 deve usare una fonte blu *large*

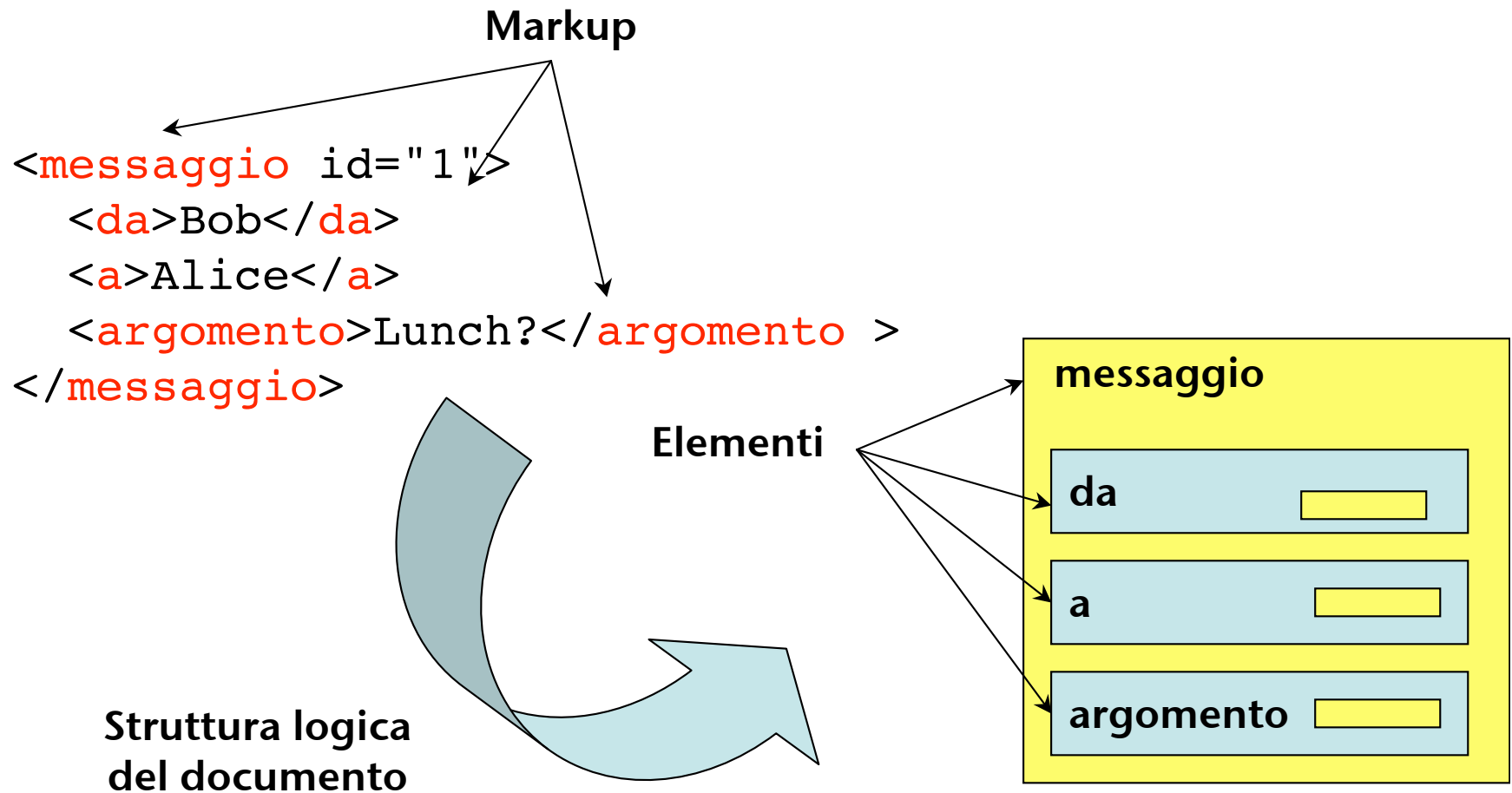
# SGML

- SGML venne inventato da IBM e nel 1986 fu standardizzato da ISO
- SGML inizialmente doveva servire a definire documenti strutturati nel campo giuridico
- E' un metalinguaggio complesso
- Sia HTML che XML derivano da SGML

# XML


- XML venne introdotto nel 1996 e standardizzato nel 1998 dal W3C come metalinguaggio di markup ("SGML per il Web")
- E' derivato da SGML, che semplifica
- Si basa sulle stesse tecnologie di base di HTML
- Elimina la debolezza di HTML (i cui tag mescolano struttura e layout), separando la manipolazione della struttura dei documenti dalle problematiche di presentazione del layout

# Esempio di frammento XML





**INTEGRATED PROTEIN INFORMATICS RESOURCE  
FOR GENOMIC AND PROTEOMIC RESEARCH**

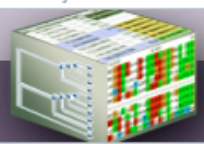


The Universal Protein Resource (UniProt) provides the scientific community with a single, centralized, authoritative resource for protein sequences and functional information.

Current release: 6.7

[UniProtKB](#) | [UniRef](#) | [UniParc](#)

**PIRSF**  
Protein Family Classification System



- Classification reflecting evolutionary relationships of full-length proteins
- [Functional site](#) and [protein name](#) rules
- [\\*Sample family report\\*](#)

**iProClass**  
Integrated Protein Knowledgebase



- Value-added reports for [UniProtKB](#) and unique [UniParc](#) proteins
- Functional analysis and [protein mapping](#)
- [\\*Sample protein report\\*](#)

**iProLINK**  
Literature, Information & Knowledge



- Source for text mining and ontology development
- [BioThesaurus](#) of protein/gene names and [RLIMS-P](#) text mining tool
- [Bibliography mapping](#)

**OTHER RESOURCE**

- [Proteomics](#): NIAID Biodefense Proteomics Admin. Center
- [PIR Grid-Enablement](#): Data node on NCI's [caBIG](#)

**P PEPTIDE SEARCH** ?

DATABASE: [iProClass](#)

Use single letter amino acid code

**T TEXT SEARCH** ?

DATABASE: [iProClass](#)

▪ [Advanced Search](#)

**\*\*Please note that this PIR beta site is under testing\*\***  
 Please send your feedback or report errors to: [✉](#)

# Descrizione di proteina in XML

```
Address http://pir.georgetown.edu/orc-cgi/parser
<?xml version="1.0" ?>
- <ProteinEntry id="JC4383">
  - <header>
    <uid>JC4383</uid>
    <accession>JC4383</accession>
    <date-added>17-Jan-1996</date-added>
    <update-seq>19-Apr-1996</update-seq>
    <update-txt>18-Jun-1999</update-txt>
  </header>
  - <protein>
    <name>3'-phosphoadenosine-5'-phosphosulfate synthetase</name>
    <alt-name>adenosine 5'-phosphosulfate kinase</alt-name>
    <alt-name>PAPS</alt-name>
    <contains>adenylylsulfate kinase (EC 2.7.1.25)</contains>
    <contains>sulfate adenylyltransferase (EC 2.7.7.4)</contains>
  </protein>
  - <organism>
    <source>spoonworm (Urechis caupo)</source>
    <formal>Urechis caupo</formal>
  </organism>
  - <reference>
    - <refinfo refid="JC4383">
      <author>Rosenthal, E.</author>
      <author>Leustek, T.</author>
      <citation>Gene 165, 243-248, 1995</citation>
      <title>A multifunctional Urechis caupo protein, PAPS synthetase,
        has both ATP sulfurylase and APS kinase activities.</title>
      <xref>MUID:96096529</xref>
    </refinfo>
    - <accinfo label="ROS">
      <accession>JC4383</accession>
      <mole_type>mRNA</mole_type>
      <seq-spec>1-610</seq-spec>
      <xref>GB:L39001</xref>
      <xref>NID:g705384</xref>
      <xref>PIDN:AAB00139.1</xref>
      <xref>PID:g705385</xref>
```

# Siti curiosi

The Million Dollar Homepage™ 1,000,000 pixels • \$1 per pixel • Own a piece of internet history! Sold: \$99,500 Available: 1,000

Get my newsletter: your@email.address Go! Homepage | Buy Pixels | FAQ | Blog | Pixel List | Press | Testimonials | Tell a friend | Contact me

Get my newsletter: your@email.address Go! Homepage | Buy Pixels | FAQ | Blog | Pixel List | Press | Testimonials | Tell a friend | Contact me

1,000,000 pixels • \$1 per pixel • Own a piece of internet history! Sold: \$99,500 Available: 1,000

Own my last 1,000 pixels - eBay auction here

Get my newsletter: your@email.address Go! Homepage | Buy Pixels | FAQ | Blog | Pixel List | Press | Testimonials | Tell a friend | Contact me



The Million Dollar Homepage © 2005 Alex Tew. All rights reserved. I am not responsible for the content of external sites. Images featured on homepage are © of their respective owners.

# Internet Archive [www.archive.org](http://www.archive.org)






Memorizza dal 1996 le pagine del Web, anche dopo che vengono eliminate dalla rete

- Un romanzo 1 MB
- *Encyclopaedia Britannica* (2,619 pagine) 1 GB
- L'antica Biblioteca di Alessandria (400,000 papiri) 800 GB
- Una biblioteca pubblica (300,000 libri) 3 TB
- Un video store (5,000 video, 1 GB/h) 8 TB
- Una stazione radio (10,000 LP/CDs, o 15,000 h, 535MB/h) 8 TB
- Intero Internet Archive al 3/2000 14 TB
- Library of Congress (20 milioni di libri) 20 TB
- Intero Internet Archive al 3/2001 43+ TB
- Intero Internet Archive al 2006 1000+ TB
- Cresce di circa 20TB al mese

# Esempio di domanda del test

- Un qualunque browser Web visualizza direttamente file con estensione
  - .doc
  - .txt 
  - .html 
  - .xls

# Esempio di domanda del test

- Quali di queste sono URL valide?
  - `cs.unibo.it` 
  - `www.cs.unibo.it` 
  - `WWW.cs.UniBo.it` 
  - `www.cs.unibo.it:81` 
  - `www.cs.unibo.it/Prova` 
  - `www.cs.unibo.it\prova`

# Esempio di domanda del test

- Cosa rappresenta la prima parte dell' URL, "http://www.unibo.it/informazioni.html"?

– Un protocollo

– Un server

– Un sito Web

– Un file





# DILBERT<sup>®</sup>

BY

SCOTT ADAMS

