

HTML perché, quando e come



Creative Commons License
Attribution-NonCommercial-ShareAlike 2.5
<http://creativecommons.org/licenses/by-nc-sa/2.5/>

claudio@cicali.org

Dalla email (testo) ai documenti (strutturati)

Il problema:

trasferire documenti strutturati attraverso Internet

Zuppa alla pavese

Ingredienti per 6 persone

Brodo bollente

Sugo di carne leggero

Fettine di burro impanato

Burro

Uova intere (freschissime)

Formaggio parmigiano

Sale

Preparazione:


Si abbrustoliscono, o meglio si friggono nel burro, a color d'oro, due fettine di pane e si accomodano in una scodella.


Su queste fette si rompono due uova intere (freschissime) che si condiscono con un po' di parmigiano e un pizzico di sale.

Nella scodella si versa il brodo bollente, piano piano, da un lato, affinché le uova non si rompano.

E' facoltativo aggiungere una mezza cucchiata di sugo di carne leggero.

Nome  **Zuppa alla pavese**

Lista degli ingredienti  **Ingredienti per 6 persone**
Brodo bollente
Sugo di carne leggero
Fettine di burro impanato
Burro
Uova intere (freschissime)
Formaggio parmigiano
Sale

Istruzioni  **Preparazione:**
Si abbrustoliscono, o meglio si friggono nel burro, a color d'oro, due fettine di pane e si accomodano in una scodella.
Su queste fette si rompono due uova intere (freschissime) che si condiscono con un po' di parmigiano e un pizzico di sale.
Nella scodella si versa il brodo bollente, piano piano, da un lato, affinché le uova non si rompano.
E' facoltativo aggiungere una mezza cucchiata di sugo di carne leggero.

Problema:

trasmettere la ricetta, compresa l'informazione delle parti che la compongono
(la sua STRUTTURA)

File con la struttura

Dalla riga 1 alla riga 1, sezione NOME

Dalla riga 3 alla riga 10, sezione INGREDIENTI

Dalla riga 12 alla riga 23, sezione ISTRUZIONI

File con la presentazione

Sezione NOME, colore rosso

Sezione INGREDIENTI, stile corsivo

Sezione qualsiasi, colore nero

File con il testo

1: Zuppa alla pavese

2:

3: Ingredienti per 6 persone

4: Brodo bollente

5: Sugo di carne leggero

6: Fettine di burro impanato

7: Burro

8: Uova intere (freschissime)

9: Formaggio parmigiano

10: Sale

11:

12: Preparazione:

13: Si abbrustoliscono, o meglio si friggono nel

14: burro, a color d'oro, due fettine di pane e si

15: accomodano in una scodella.

16: Su queste fette si rompono due uova intere

17: (freschissime) che si condiscono con un po' di

18: parmigiano e un pizzico di sale.

19: Nella scodella si versa il brodo bollente, piano

20: piano, da un lato, affinché le uova non si

21: rompano.

22: E' facoltativo aggiungere una mezza cucchiata

23: di sugo di carne leggero.

Problemi della soluzione adottata

1. Chi ha scelto la convenzione da adottare per la struttura?
2. Possiamo solo trasmettere ricette!
3. Altra ricetta, altro file struttura
4. Tre file sono troppi: se uno non viene trasmesso o perso?

La soluzione stava in un sistema che includesse tutte le informazioni all'interno di un unico file: **struttura + dati**.



SGML

il linguaggio per scrivere linguaggi markup

METALINGUAGGIO

ISO 8879:1986 Information processing—Text and office systems—**Standard Generalized Markup Language** (SGML)

Markup: dati + informazioni aggiuntive

Tag & Elementi

Elemento **PINCOPALLO**

Tag & Elementi

Elemento **PINCOPALLO**

<PINCOPALLO>

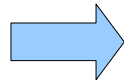
Contenuto dell'elemento

</PINCOPALLO>

Tag & Elementi

Elemento **PINCOPALLO**

TAG di apertura



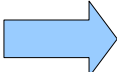
<PINCOPALLO>

Contenuto dell'elemento


</PINCOPALLO>

Tag & Elementi

Elemento **PINCOPALLO**

TAG di apertura  **<PINCOPALLO>**

Contenuto dell'elemento

TAG di chiusura  **</PINCOPALLO>**

Tag & Elementi

Elementi VUOTI (non hanno tag di chiusura)

```
<ELEMENTO_VUOTO />
```

Esiste un rapporto **gerarchico** tra gli elementi

e

Deve esistere un nodo **radice** che contenga tutti gli altri

<pincopallo>

Testo testo testo

</pincopallo>

<pincopallo>

Testo testo testo

<supermario>

Altro testo testo testo

</supermario>

</pincopallo>

<radice>

<pincopallo>

Testo testo testo

<supermario>

Altro testo testo testo

</supermario>

</pincopallo>

</radice>

L'indentazione aumenta la leggibilità

```
<radice>
```

```
  <pincopallo>
```

```
    Testo testo testo
```

```
      <supermario>
```

```
        Altro testo testo testo
```

```
      </supermario>
```

```
    </pincopallo>
```

```
</radice>
```

I tag non possono “incrociarsi”

```
<radice>
```

```
  <pincopallo>
```

```
    Testo testo testo
```

```
      <supermario>
```

```
        Altro testo testo testo
```

```
      <pincopallo>
```

```
        </supermario>
```

```
</radice>
```

<RICETTA>

<NOME>

Zuppa alla pavese

</NOME>

<INGREDIENTI>

Ingredienti per **<PERSONE>6</PERSONE>** persone

Brodo bollente

Sugo di carne leggero

Fettine di burro impanato

Burro

Uova intere (freschissime)

Formaggio parmigiano

Sale

</INGREDIENTI>

<PREPARAZIONE>

Si abbrustoliscono, o meglio si friggono nel burro, a color d'oro, due fettine di pane e si accomodano in una scodella. Su queste fette si rompono due uova intere (freschissime) che si condiscono con un po' di parmigiano e un pizzico di sale. E' facoltativo aggiungere una mezza cucchiata di sugo di carne leggero.

</PREPARAZIONE>

</RICETTA>

Gli attributi

<elemento **attributo="valore"**>

Gli attributi

```
<elemento attributo="zero"  
attributo_1="uno"  
attributo_2=due  
attributo_3=tre>
```

```
<RICETTA LINGUAGGIO="IT" DIFFICOLTA="3">  
  <NOME>Zuppa alla pavese</NOME>  
  <INGREDIENTI PERSONE="6">
```

...

```
</RICETTA>
```

Dal SGML all'HTML

Problema: potenzialmente, esistevano decine di tipi diversi di documenti

Dal SGML all'HTML

Problema: potenzialmente, esistevano decine di tipi diversi di documenti

Soluzione: un linguaggio markup quanto più generico possibile

1989: nasce un linguaggio standard per la struttura dei documenti usufruiti sul web

Ma in che senso è uno STANDARD?





- Fondato nel 1994
- Definisce gli STANDARD per il WEB
- E' *super partes, vendor neutral*
- > 270 organizzazioni membre
- Il direttore è Tim Berners-lee (inventore del web)
- Mette a disposizione tool e tecnologie per gli sviluppatori

W3C operations are jointly administered by the **MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)** in the USA, the **European Research Consortium for Informatics and Mathematics (ERCIM)** headquartered in France and **Keio University** in Japan.
W3C also has World Offices in fifteen regions around the world.

L'Ufficio Italiano W3C è ospitato dall'
Istituto di [Scienza e Tecnologie dell' Informazione "A. Faedo"](#) (ISTI) del [C.N.R.](#) a Pisa.



Genesi di uno “standard”:

- ***Submission*** quando vi è una proposta di standardizzare un linguaggio, una procedura, un formato...
- ***Working Draft*** quando la proposta è accettata e si inizia a produrre della documentazione
- ***Proposed Recommendation*** prima (o successive) versioni finali della raccomandazione
- ***Recommendation*** la versione finale di una versione dello standard

<http://www.w3.org/TR/>



“Mi raccomando!”

HTML 0.9, 1989

il linguaggio di HTML disponibile sul browser WWW aveva strutture di base per intestazioni, paragrafi e stili base, più ovviamente il tag A che permetteva di effettuare “link” tra un documento e un altro, che ne costituiva la novità.

HTML 1.0, 1992

I primi browser shareware e freeware al mondo implementavano alcune versioni di HTML leggermente diverse. Tra queste ebbe un certo successo la proposta HTML+. Viene introdotto il tag IMG e il supporto per il formato grafico GIF.

HTML 2.0, 1994

La prima versione veramente nota di HTML. Questa è quella implementata su Mosaic, da cui deriverà Netscape. E' la prima versione ad essere formalizzata su un DTD SGML, invece di ispirarsi vagamente ad SGML. Introduce i form per l'inserimento dati da parte dell'utente.

HTML 3.2, 1997

Quando divenne chiaro che i browser non avrebbero supportato tutto il 3.0, si lavorò per generarne un sottoinsieme su cui ci fosse consenso e che tenesse conto delle aggiunte proprietarie dei vari produttori. Questa versione include tabelle, applet, script e altre migliorie, ma non i frame, sebbene Netscape e Microsoft le avessero già implementate fin dal 1995.

HTML 4.0, 1997

supporto per l'internazionalizzazione, per gli style sheet, per i frame, tabelle molto più ricche, il tag OBJECT, ecc.

HTML 4.01, 1999

contiene alcune minime variazioni e correzioni

1993 NCSA Mosaic 1.0
1994 Netscape 1.0
1995 Microsoft Windows IE 1.0
1996 Microsoft Windows IE 2.0, 3.0, Netscape 2.0 e 3.0, Opera 2.1
1997 Microsoft Windows IE 4.0, Netscape 4.0, Opera 3.0
1998 Netscape 4.5, Opera 3.5. Progetto Mozilla
1999 Microsoft Windows IE 5.0
2000 Microsoft Windows IE 5.5
2001 Microsoft Windows IE 6.0, Opera 6.0
2002 Netscape 7.0, Mozilla 1.0
2003 Opera 7.x, Apple Safari
2004 Firefox 1.0
2005 Firefox 1.5, Safari 2, Opera 9
2006 Firefox 2.0, Safari 3
2007 IE7 ...

Le versioni dei browser, le versioni di HTML e tutto il resto...

DTD



Document Type Definition



Document Type Definition

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
```

La primissima riga del documento!

HTML

- Non è un linguaggio di programmazione
- E' un'applicazione SGML
- Non è *case sensitive*
- I documenti HTML sono file di testo
- Il tipo di *fine riga* non ha importanza
- L'estensione è .html o .htm (ma in teoria non ha importanza)

HTML: I commenti

```
<!-- Questo è un commento -->
```


```
<!-- Questo  
è su linee  
multiple  
-->
```


HTML: La struttura di base

Ogni documento HTML è suddivisibile in tre parti principali

1. DOCTYPE
2. TESTATA
3. CORPO

HTML: La struttura di base

Document type  `<!DOCTYPE ... >`

Elemento root (html)  `<HTML>`

Testata (head)  `<HEAD>`

...

`</HEAD>`

Corpo (body)  `<BODY>`

...

`</BODY>`

`</HTML>`

Elemento HEAD

HEAD

TITLE ← obbligatorio!

META

BASE

Elemento HEAD

HEAD

TITLE

META name="nome del meta" content="una stringa"

BASE

Elemento HEAD

HEAD

TITLE

META

BASE href="url"

<http://diveintomark.org/archives/2004/01/02/relative-uris>

Elemento HEAD

HEAD

TITLE

META

BASE

STYLE
SCRIPT
LINK

Elementi per il testo

P

PRE

EM

STRONG

CITE

Q

BLOCKQUOTE

Liste

UL

LI

OL

LI

DL

DT

DD

Una distinzione fondamentale:

Elementi **TEXT LEVEL** (inline) e **BLOCK LEVEL**

Titoli

H1 H2 H3 H4 H5 H6

Attributi comuni a tutti gli elementi

ID

CLASS

```
<h1 id="TitoloDocumento">  
  La zanzara anofele  
</h1>
```

```
<ul class="menuSezione">  
  ...  
</ul>
```

Non possono iniziare con un numero

Devono iniziare con una lettera e seguiti da qualsiasi numero di lettere, cifre, '-', '_', ':' e '.').

Altri attributi comuni

ID

CLASS

LANG

DIR

STYLE

TITLE

Elementi strutturali generici

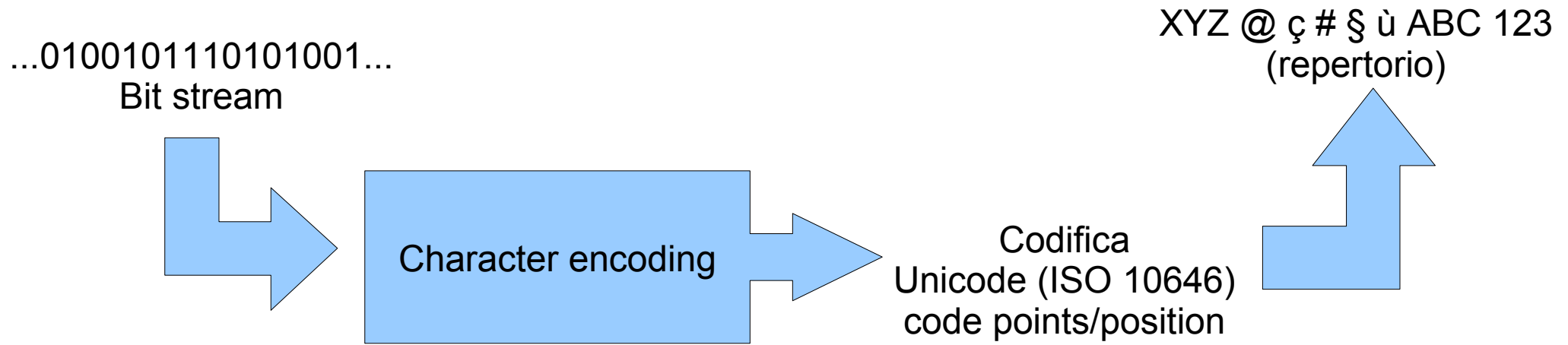
DIV

SPAN

Character set / Character encoding: la tabella ASCII

000	NUL	033	!	066	B	099	c	132	ä	165	Ñ	198	ã	231	b
001	Start Of Header(SOH)	034	"	067	C	100	d	133	å	166	ª	199	Ä	232	ß
002	Start Of Text (STX)	035	#	068	D	101	e	134	ä	167	º	200	ℒ	233	Ù
003	End Of Text (ETX)	036	\$	069	E	102	f	135	ç	168	¿	201	ℓ	234	Ú
004	End Of Transmission (EOT)	037	%	070	F	103	g	136	è	169	®	202	ℓ	235	Û
005	Enquiry	038	&	071	G	104	h	137	ë	170	¬	203	π	236	Ý
006	Acknowledge (ACK)	039		072	H	105	i	138	è	171	½	204	⌈	237	Ÿ
007	Bell	040	(073	I	106	j	139	ï	172	¼	205	=	238	ˉ
008	Backspace (BS)	041)	074	J	107	k	140	î	173	ı	206	‡	239	˘
009	Horizontal Tab	042	*	075	K	108	l	141	ì	174	«	207	×	240	-
010	Line Feed (LF)	043	+	076	L	109	m	142	Å	175	»	208	ø	241	±
011	Vertical Tab	044	,	077	M	110	n	143	Ä	176	∴	209	ð	242	_
012	Form Feed (FF)	045	-	078	N	111	o	144	É	177	∑	210	É	243	¼
013	Carriage Return (CR)	046	.	079	O	112	p	145	æ	178	⊠	211	Ê	244	⌈
014	Shift Out	047	/	080	P	113	q	146	Æ	179		212	Ë	245	§
015	Shift In	048	0	081	Q	114	r	147	ó	180	¡	213	ì	246	÷
016	Dataline Escape (DLE)	049	1	082	R	115	s	148	ö	181	À	214	í	247	˘
017	DC 1 (XON)	050	2	083	S	116	t	149	ò	182	Á	215	î	248	ª
018	DC 2	051	3	084	T	117	u	150	ù	183	Â	216	ï	249	˘
019	DC 3 (XOFF)	052	4	085	U	118	v	151	ú	184	⊙	217	ª	250	.
020	DC 4	053	5	086	V	119	w	152	ÿ	185	‡	218	ª	251	'
021	Negative Acknowledge (NAK)	054	6	087	W	120	x	153	Ö	186		219	■	252	˘
022	Synchronous Idle	055	7	088	X	121	y	154	Ü	187	⌈	220	■	253	˘
023	End Of Transmission Block	056	8	089	Y	122	z	155	ø	188	⌋	221	ı	254	■
024	Cancel	057	9	090	Z	123	{	156	£	189	¢	222	ı	255	
025	End Of Medium	058	:	091	[124		157	∅	190	¥	223	■		
026	Substitute	059	;	092	\	125	}	158	×	191	⌈	224	Ó		
027	Escape (ESC)	060	<	093]	126	~	159	f	192	ℒ	225	ß		
028	File Separator	061	=	094	^	127 (DEL)	␣	160	á	193	⌊	226	Ö		
029	Group Separator	062	>	095	_	128	Ç	161	í	194	⌈	227	Ö		
030	Record Separator	063	?	096	`	129	ú	162	ó	195	⌈	228	ø		
031	Unit Separator	064	@	097	a	130	é	163	ú	196	-	229	Ö		
032	SPACE (SP)	065	A	098	b	131	â	164	ñ	197	†	230	μ		

Character set / Character encoding



Come si specifica il character encoding di un documento?

Come si specifica il character encoding di un documento?

1. Lo dichiara il server HTTP
2. Lo si specifica con un META
3. Attributo charset

Questo è anche l'ordine di rilevanza

Character references

&#nnn;

&#xnnn;

0 rappresenta '0'

A rappresenta 'A'

< rappresenta '<'

> rappresenta '>'

Entity references (o named)

&entity_name;

< rappresenta '<'

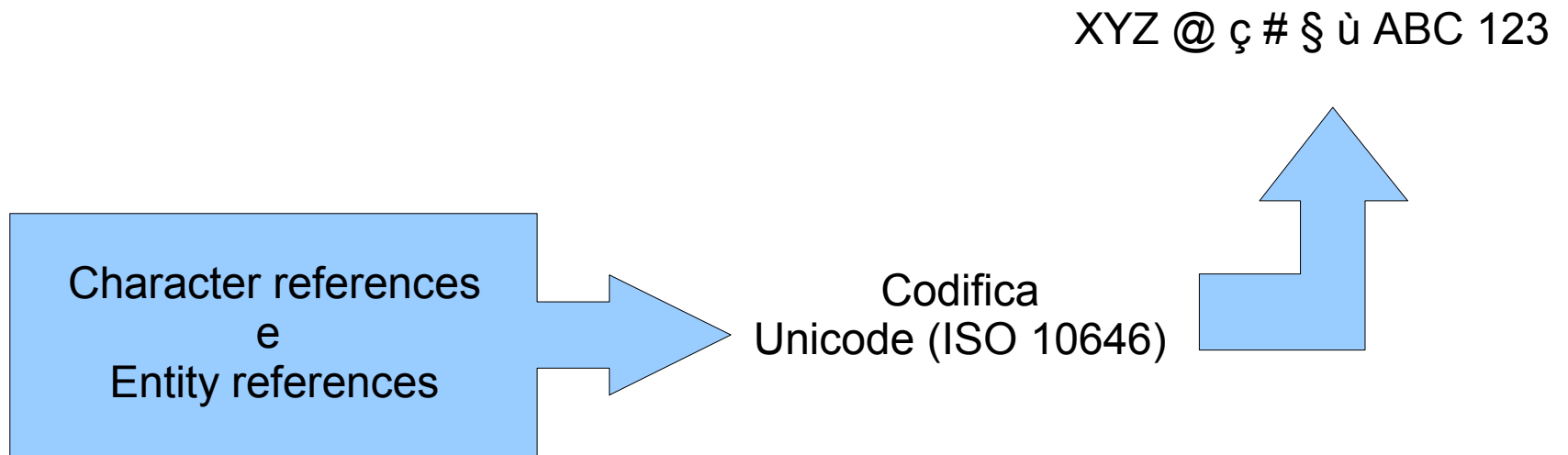
> rappresenta '>'

è rappresenta 'è'

é rappresenta 'é'

Character set / Character encoding

Character & Entity references





€

&x20ac;

€

Gestione degli spazi

Le tabelle

```
<table summary="Lista clienti">  
  <caption>Lista clienti</caption>
```

```
</table>
```

Le tabelle

```
<table>  
  <caption>Lista clienti</caption>  
  <thead>  
</thead>  
  <tfoot>  
</tfoot>  
  <tbody>  
</tbody>  
  
</table>
```

Le tabella

```
<table>
  <caption>Lista clienti</caption>
  <thead>
    <tr>
      <th>Nome</th>
      <th>Cognome</th>
    </tr>
  </thead>
  <tfoot>
  </tfoot>
  <tbody>
  </tbody>
</table>
```

Le tabelle

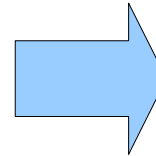
```
<table>
  <caption>Lista clienti</caption>
  <thead>
</thead>
  <tfoot>
</tfoot>
  <tbody>
    <tr>
      <td>Claudio</td>
      <td>Cicali</td>
    </tr>
  </tbody>
</table>
```

Le tabelle

```
<table>
  <caption>Lista clienti</caption>
  <thead>
</thead>
  <tfoot>
    <tr>
      <td>Totale clienti</td>
      <td>13</td>
    </tr>
  </tfoot>
  <tbody>
</tbody>
</table>
```

Le tabelle

```
<table summary="Lista clienti di flexer">
  <caption>Lista clienti</caption>
  <thead>
    <tr><th>Nome</th><th>Cognome</th></tr>
  </thead>
  <tfoot>
    <tr>
      <td>Totale clienti</td>
      <td>13</td>
    </tr>
  </tfoot>
  <tbody>
    <tr><td>Claudio</td>
      <td>Cicali</td>
    </tr>
  </tbody>
</table>
```



Lista clienti

Nome	Cognome
Claudio	Cicali
Totale	13

Le tabelle: raggruppare le colonne

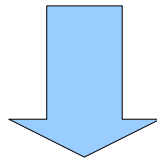
```
<table>  
  <colgroup>  
    <col>  
    <col>  
  </colgroup>  
  <thead>  
  . . . .  
</table>
```

Le tabelle: raggruppare le colonne

```
<table>  
  <colgroup>  
    <col width="30">  
    <col width="0*">  
    <col width="1*">  
    <col width="2*">  
    <col width="3*">  
  </colgroup>  
  <thead>  
  ....  
</table>
```

Le tabelle: unire celle e righe

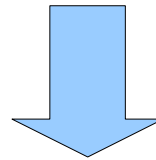
```
<table>
  <tr>
    <td>Salve</td>
    <td>Hello</td>
  </tr>
  <tr>
    <td colspan="2">Saludos</td>
  </tr>
</table>
```



Salve	Hello
Saludos	

Le tabelle: unire celle e righe

```
<table>  
  <tr>  
    <td>Salve</td>  
    <td>Hello</td>  
    <td rowspan="2">Azione</td>  
  </tr>  
  <tr>  
    <td colspan="2">Saludos</td>  
  </tr>  
</table>
```



Salve	Hello	Azione
Saludos		

1 + 2		3
4	5	6
7 + 8 + 9		

```
<table>
  <tr>
    <td colspan="2">1+2</td>
    <td>3</td>
  </tr>
  <tr>
    <td>4</td>
    <td>5</td>
    <td>6</td>
  </tr>
  <tr>
    <td colspan="3">7 + 8 + 9</td>
  </tr>
</table>
```

1	2		3
4	5	6	
	7		

1	2		3
4	5	6	
	7		

```
<table>
  <tr>
    <td>1</td>
    <td colspan="2">2</td>
    <td rowspan="3">3</td>
  </tr>
  <tr>
    <td rowspan="2">4</td>
    <td>5</td>
    <td>6</td>
  </tr>
  <tr>
    <td colspan="2">7</td>
  </tr>
</table>
```

URL

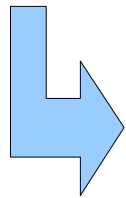
(Uniform Resource Locator)

`http://www.example.com/index.html`

URL

(Uniform Resource Locator)

http: //www.example.com/index.html



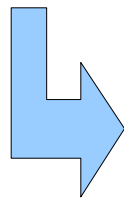
Uno *schema* (ftp, http, rtsp, mailto...)

E' seguito da un “.”

URL

(Uniform Resource Locator)

`http://www.example.com/index.html`



Un hostname

Inizia con //

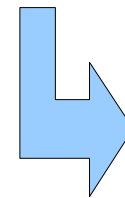
Può iniziare con *user:password@*

Può contenere l'indicazione della porta

URL

(Uniform Resource Locator)

`http://www.example.com/index.html`

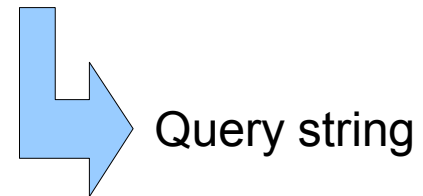


Un percorso (*path*)

URL

(Uniform Resource Locator)

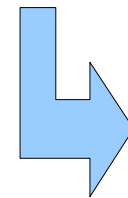
`http://www.example.com/index.html?para=1&parb=3`



URL

(Uniform Resource Locator)

`http://www.example.com/index.html#chi_siamo`



Fragment

Elemento A

1. link
2. visited
3. active

Elemento A

<p>

Documento all'interno del sito

Cerca con Google

Torna all'indice

</p>

Le immagini: l'elemento IMG

```

```

Le immagini: i formati

GIF

- 256 colori
- Trasparenza
- Animazione
- Lossless

JPEG/JPG

- Milioni di colori
- No trasparenza
- No animazione
- Lossy

PNG

- Milioni di colori
- Trasparenza (alpha channel)
- No animazione
- Lossless