

# HTML

# the language for the web

The basics

# Introduction

- HTML is VERY easy to pick up.
- There are many good books as well excellent resources on-line
- Our goal is to teach you the basics so that you can pick up the rest on your own.

# Readings

- [https://developer.mozilla.org/en-US/docs/Learn/HTML/Introduction\\_to\\_HTML](https://developer.mozilla.org/en-US/docs/Learn/HTML/Introduction_to_HTML)
- <https://www.w3schools.com/html/>
- [http://www.cellbiol.com/bioinformatics\\_web\\_development/chapter-3-your-first-web-page-learning-html-and-css/](http://www.cellbiol.com/bioinformatics_web_development/chapter-3-your-first-web-page-learning-html-and-css/)

# What is HTML?

## Hypertext Markup Language

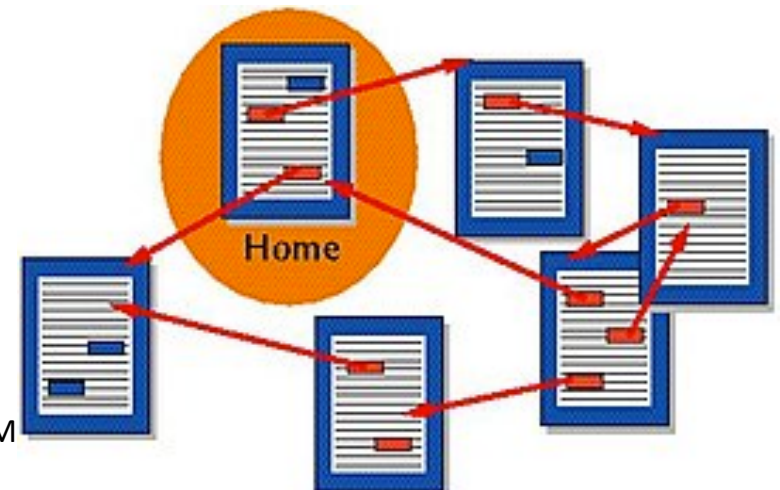
a simple data format used to create  
**hypertext** documents that are  
portable from one platform to  
another.



# Hypertext

**Hypertext** is text displayed on a computer display or other electronic devices with references (**hyperlinks**) to other text that the reader can immediately access.

Hypertext documents are interconnected by hyperlinks, which are **typically activated by user actions**, such as a mouse click, keypress set or by touching the screen.



# Markup language

Not a programming language!

A **markup language** is a computer language that uses tags to **define elements** within a document.

It provides **annotations (metadata)**.

It is **human-readable**.

Several markup languages exist, the two most popular are **HTML** and **XML** (both incarnations of **SGML**).

# Example of Markup Language

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
```

```
<data>
```

```
<NETWORK>
```

```
<IP>172.150.1.101</IP>
```

```
<IP_LODESERVER>172.150.1.3</IP_LODESERVER>
```

```
</NETWORK>
```

```
<LECTURE id="27">
```

```
<COURSE_NAME>Web Programming</COURSE_NAME>
```

```
<LECTURE_NAME>Introduction to HML</LECTURE_NAME>
```

```
<TEACHER_NAME>Marco Ronchetti</TEACHER_NAME>
```

```
<TIME>5225.00</TIME>
```

```
</LECTURE>
```

```
</data>
```

# Element, tags, content, attribute

`<chap number="1">Text for Chapter 1</chap>`

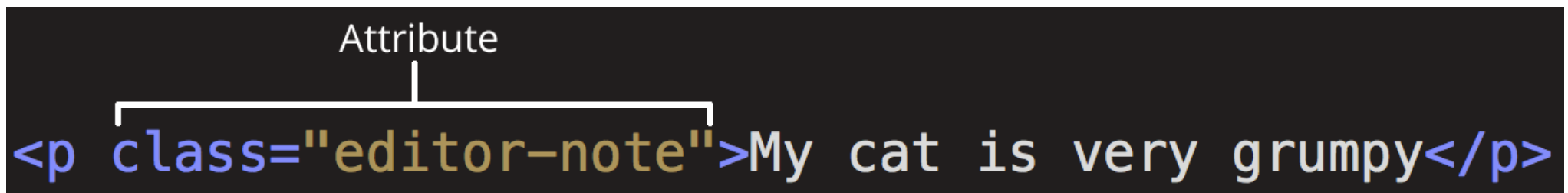
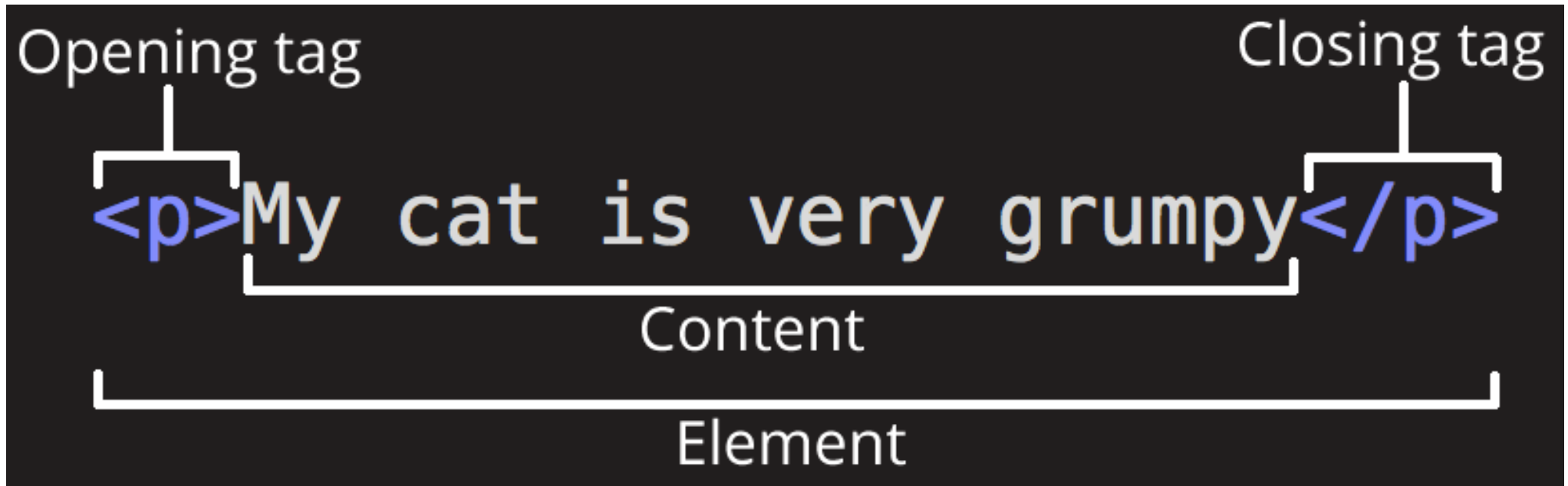
An **ELEMENT** is a sequence of characters that begins with a *start tag* and ends with an *end tag* and includes everything in between.

The characters in between the tags (rendered in **green** above) constitute the **CONTENT**.

The *start tag* may contain optional **ATTRIBUTES**.

The characters rendered in **blue** in the above element constitute an attribute. In this example, a single *attribute* provides the number value for the chapter.

# Example



# Versions

version	Date	Notes
HTML	1991	
HTML 2.0	1995	First official version
HTML 3.2	1997	
HTML 4.0 (4.01)	1999 (2001)	
XHTML 1.0	2000 (2001)	XML version of HTML 4
DHTML		HTML 4 +JS + CSS + DOM
XHTML 2.0		Abandoned 2009
HTML 5 (5.2)	2014 (2017)	(2008 first draft)
XHTML 5.1		XML version of HTML 5

Initially maintained by W3C

Since 2012 also defined by WHATWG (Web Hypertext Application Technology Working Group)

Since 2019 maintained only by WHATWG

# Most desktop browsers still miss about 10% of features!

Maximum points: 555

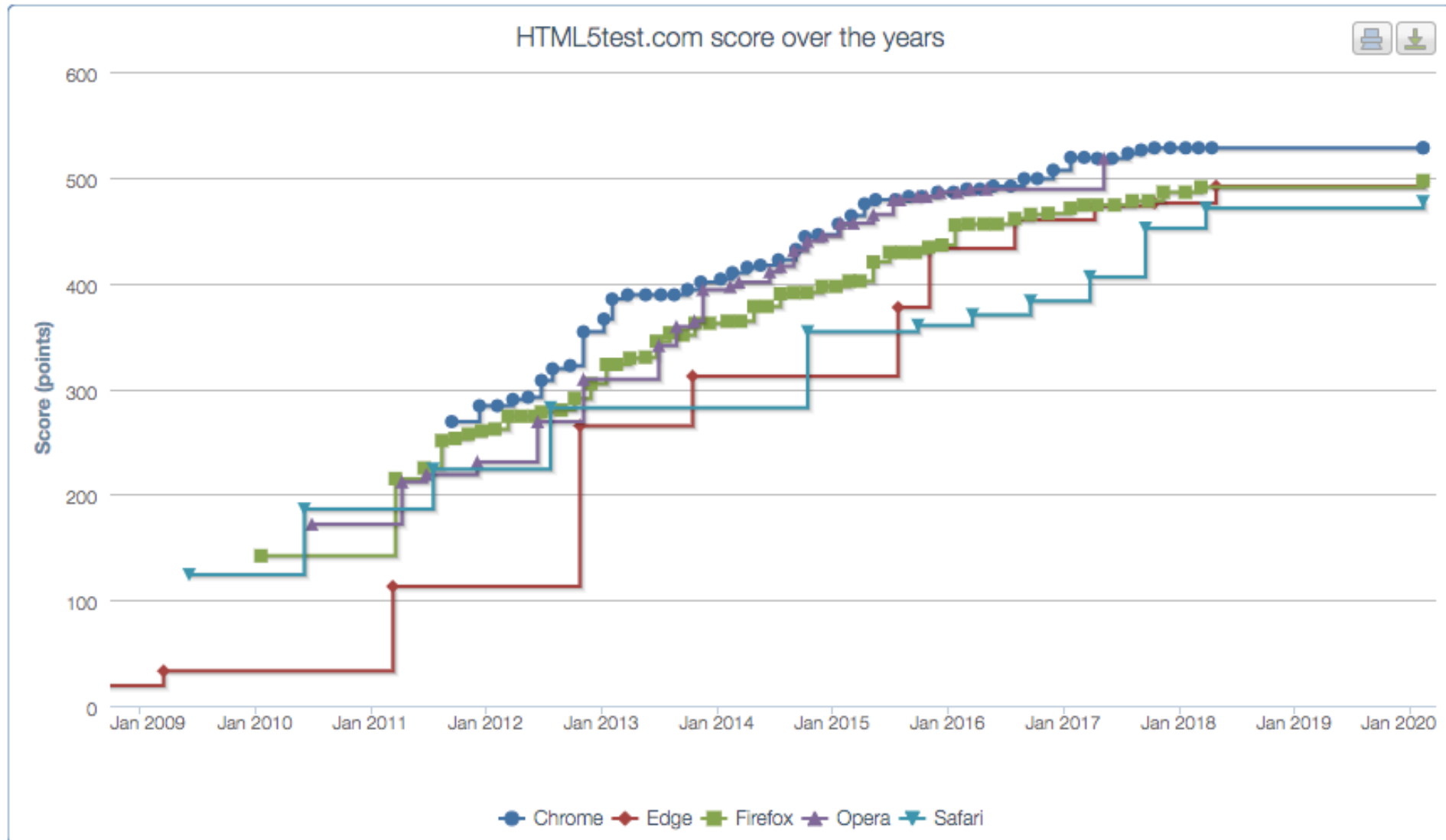
<http://html5test.com/results/desktop.html>

	Chrome	Opera	Firefox	Edge	Safari
Upcoming	67 528		60 497	18 496	11.2 477
Current	66 528	45 518	59 491	17 492	11.1 471
Older	65 528	37 489	58 486	16 476	11 452
	64 528	30 479	57 486	15 473	10.1 406
	63 528	12.10 309	56 478	14 460	10.0 383
	62 528		55 478	13 433	9.1 370
	61 526		54 474	12 377	9.0 360
	60 523		53 474	Internet Explorer	8.0 354
				11 312	

Mobile and Tablets are even worse...

# HTML5 over the years

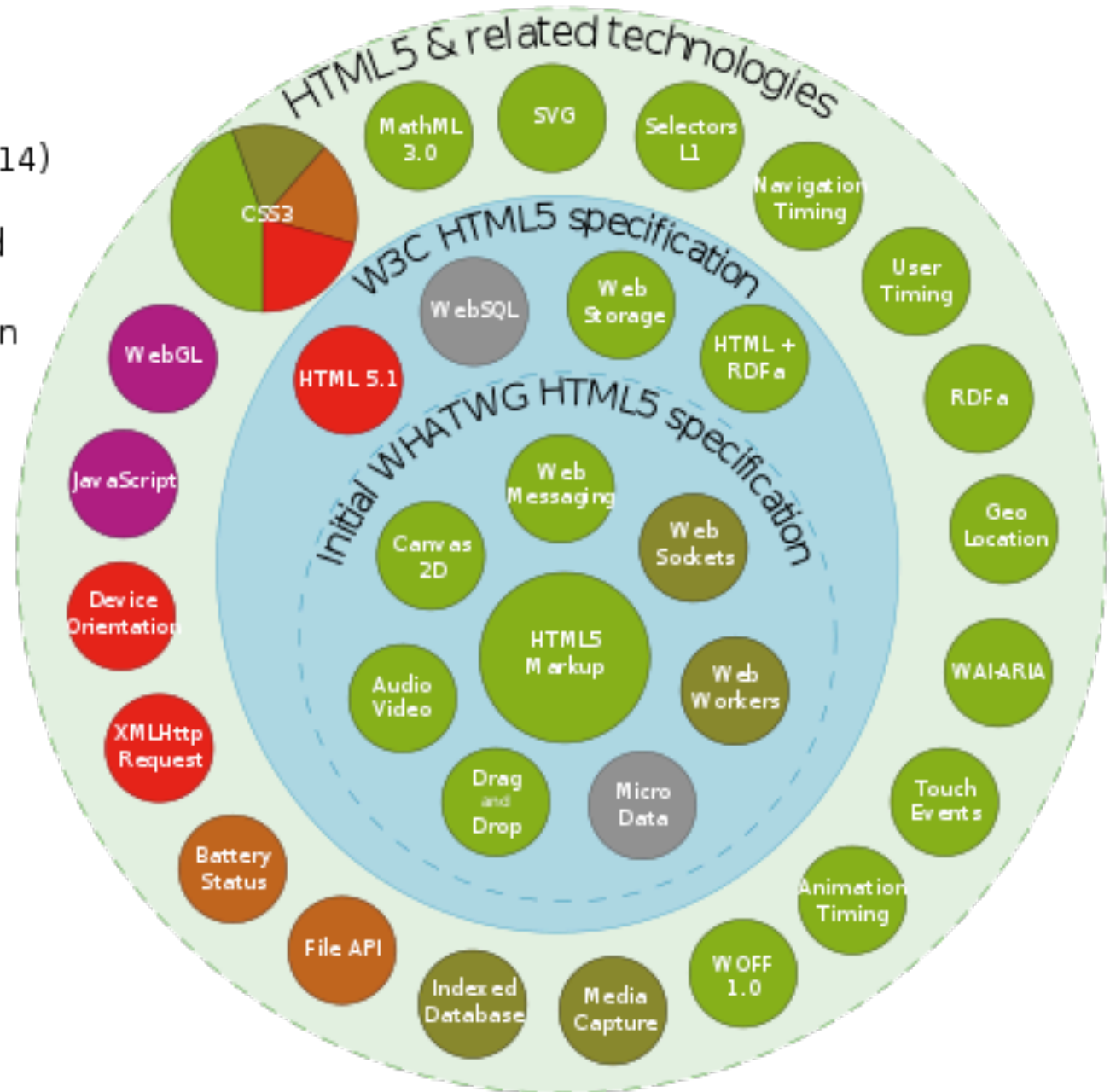
<http://html5test.com/results/desktop.html>



# HTML5

Taxonomy & Status (October 2014)

- Recommendation/Proposed
- Candidate Recommendation
- Last Call
- Working Draft
- Non-W3C Specifications
- Deprecated or inactive



Da wikimedia

# A step back: the web...

The web was born as a

- distributed document delivery system
- Using a client-server model
- Based on the notion of Hypertext
- With HTML as the *main* presentation language (markup-based)

Document sharing

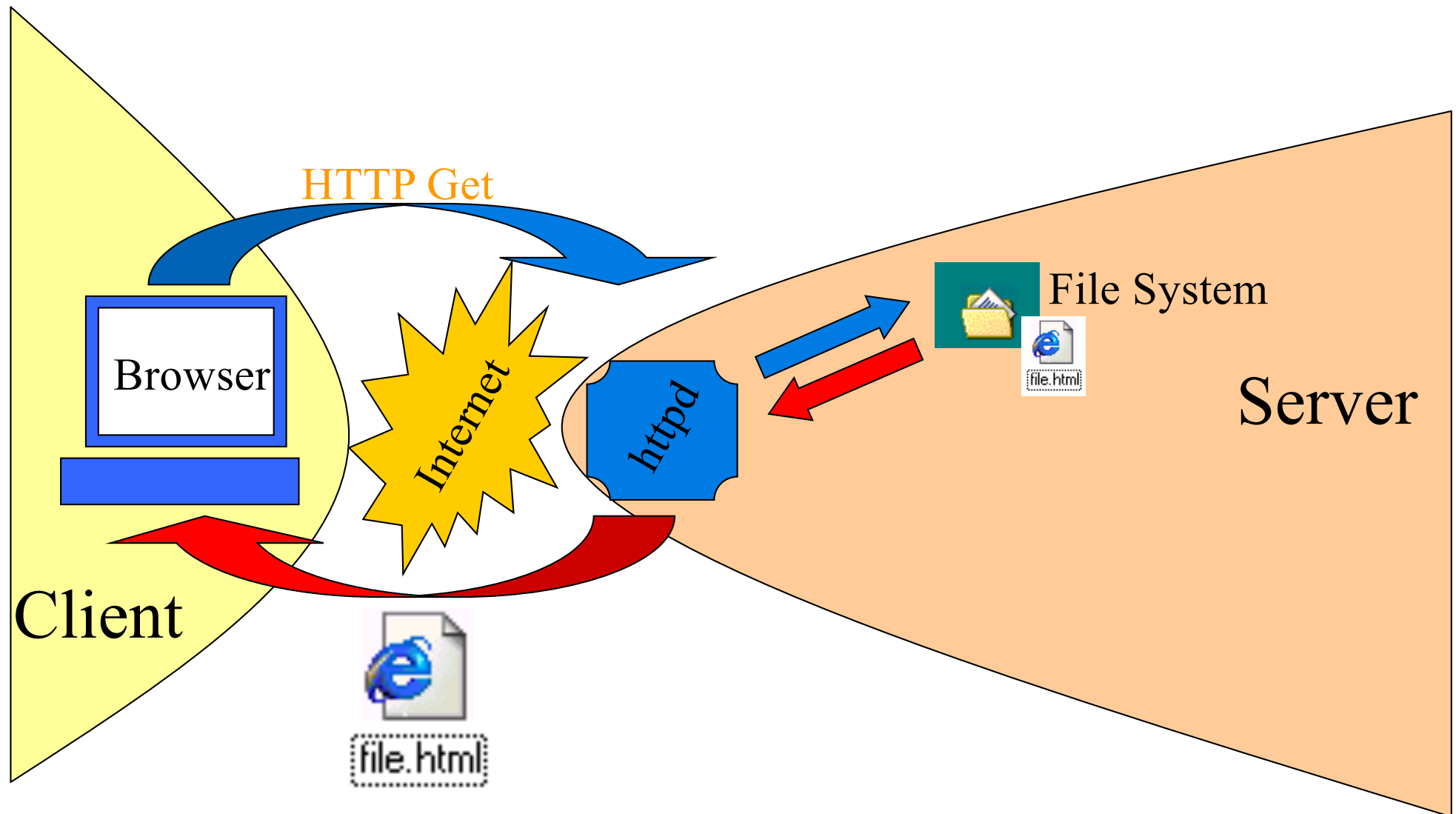
Without

- the possibility of triggering processes

No on-line shopping,  
No e-banking,  
No social media...



## The original web architecture: static pages



**Initial idea: get (static) interlinked documents**

# What is HTML?

- Defined in RFC 1866
- Logical markup language
- maximally portable
- **Graceful degradation of presentation**

“A philosophical rule was that **HTML should convey the structure of a hypertext document, but not the details of its presentation.**”

Since then, HTML has **evolved into a presentation language.**

HTML standards are developed under the authority of the World **Wide Web Consortium (W3C)**, headed by Tim Lee

<http://www.w3c.org>



# Media - audio (HTML 5)

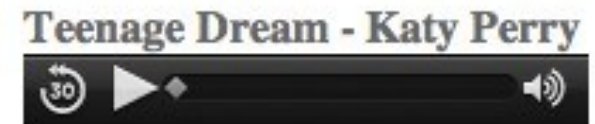
```
<audio src="teenage_dream.mp3"></audio>
```

```
<audio controls>
```

```
<source src="teenage_dream.mp3"/>
```

```
<source src="teenage_dream.ogg"/>
```

```
</audio>
```



Provides a download link for non-supporting browsers:

```
<audio src="teenage_dream.ogg" autoplay controls loop>
```

```
<a href="teenage_dream.ogg">download</a>
```

```
</audio>
```

***“Graceful degradation of presentation”***

# Media - video (HTML 5)

```
<video src="ironman.ogg" />
```

Automatically show native controls

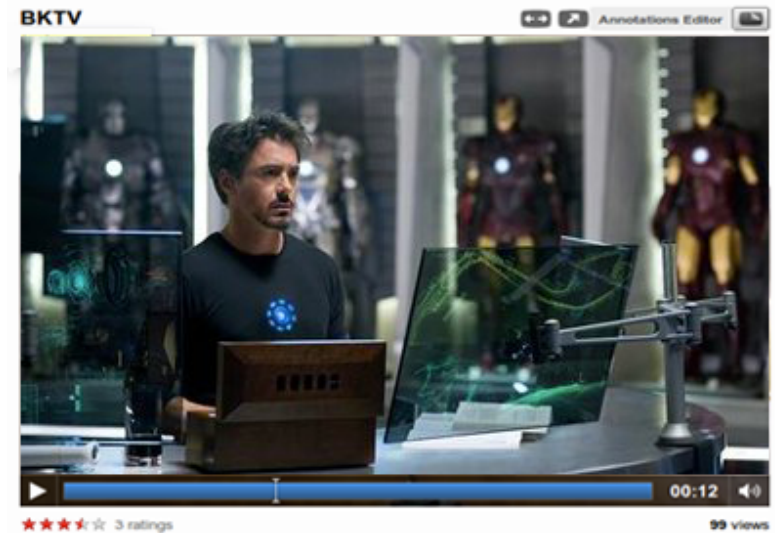
```
<video src="ironman.ogg" controls />
```

```
<video controls/>
```

```
<source src="ironman.mp4" />
```

```
<source src="ironman.ogg" />
```

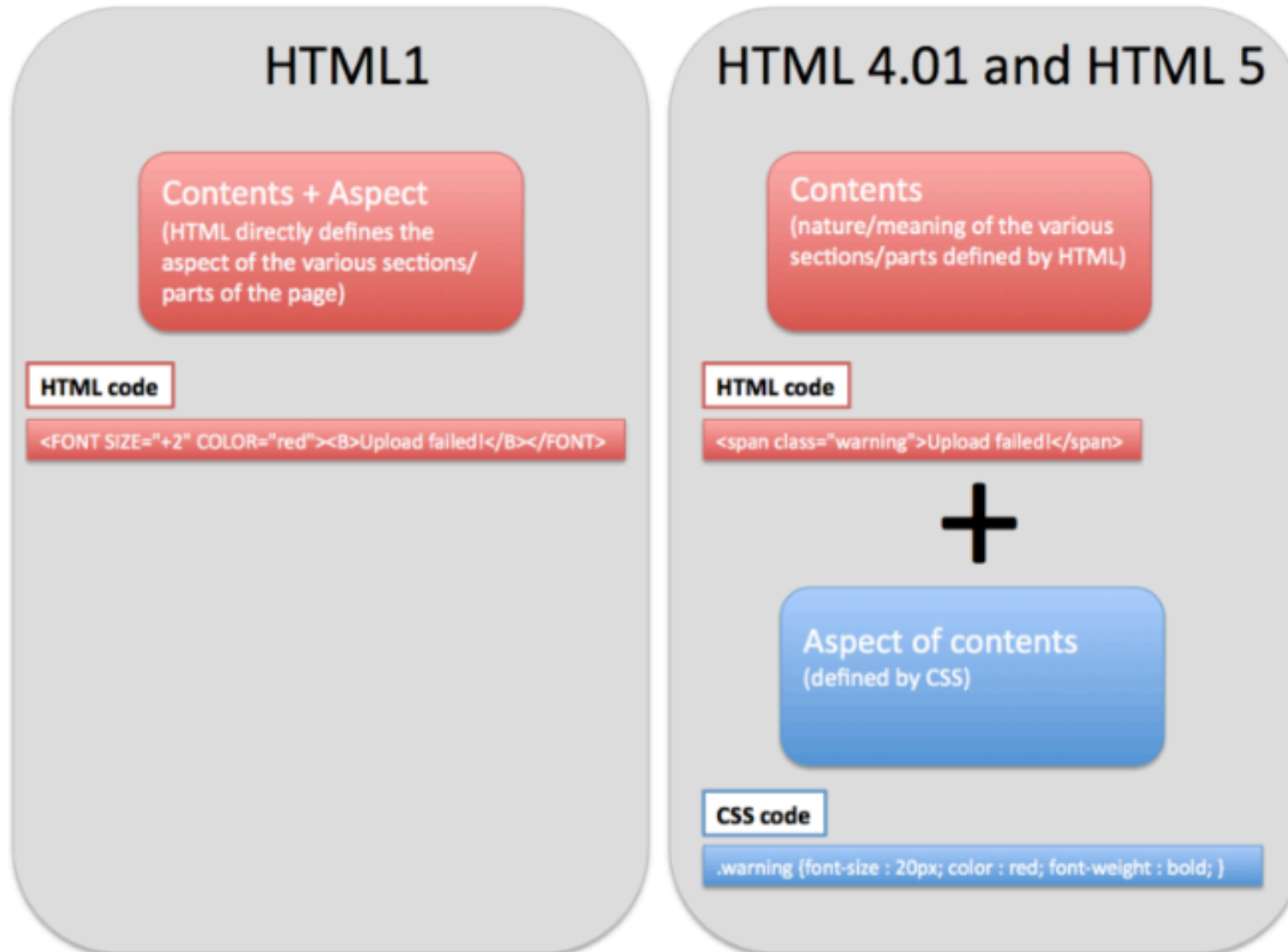
```
</video>
```



Specify multiple source elements to support more browsers

(i.e. mp4 will work in Safari, ogg will work in Firefox)

# Content/Aspect separation



This image is part of the Bioinformatics Web Development tutorial at: [http://www.cellbiol.com/bioinformatics\\_web\\_development/](http://www.cellbiol.com/bioinformatics_web_development/) © cellbiol.com, all rights reserved

Da [http://www.cellbiol.com/bioinformatics\\_web\\_development](http://www.cellbiol.com/bioinformatics_web_development)

# What is HTML?

HTML is an application of ISO Standard 8879:1986:

"Information Processing Text and Office Systems; Standard Generalized Markup Language" (**SGML**).

The HTML Document Type Definition (DTD) is a formal definition of the HTML syntax in terms of SGML.



# What is SGML?

**SGML** (Standard Generalized Markup Language) is a standard for how to specify a document markup language or tag set.

Such a specification is itself a "grammar":  
**document type definition (DTD)**

**SGML** is not in itself a document language, but a description of how to specify one.



# Structure of an HTML doc

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<title>My test page</title>
```

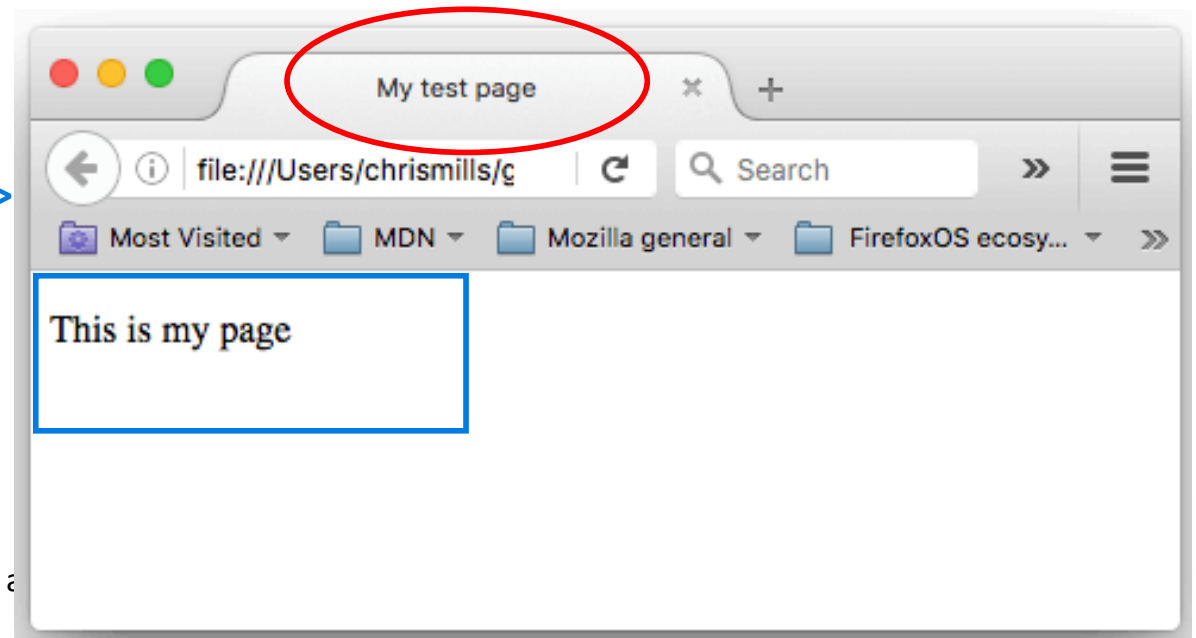
```
</head>
```

```
<body>
```

```
<p>This is my page</p>
```

```
</body>
```

```
</html>
```



# Structure of an HTML doc

- HEAD: contains information about the document (metadata):
  - Title of the page (which appears at the top of the browser window)
  - Meta tags: used to describe the content (used by Search engines)
  - Statements or references to JavaScript and StyleSheets code
- BODY: Contains the actual content of the document
  - This is the part that will be displayed in the browser window



# DocType declaration

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0  
Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-  
transitional.dtd">
```

Today (simplified in HTML 5.0):

```
<!DOCTYPE html>
```

# Structure of an HTML doc

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<title>My test page</title>
```

```
</head>
```

```
<body>
```

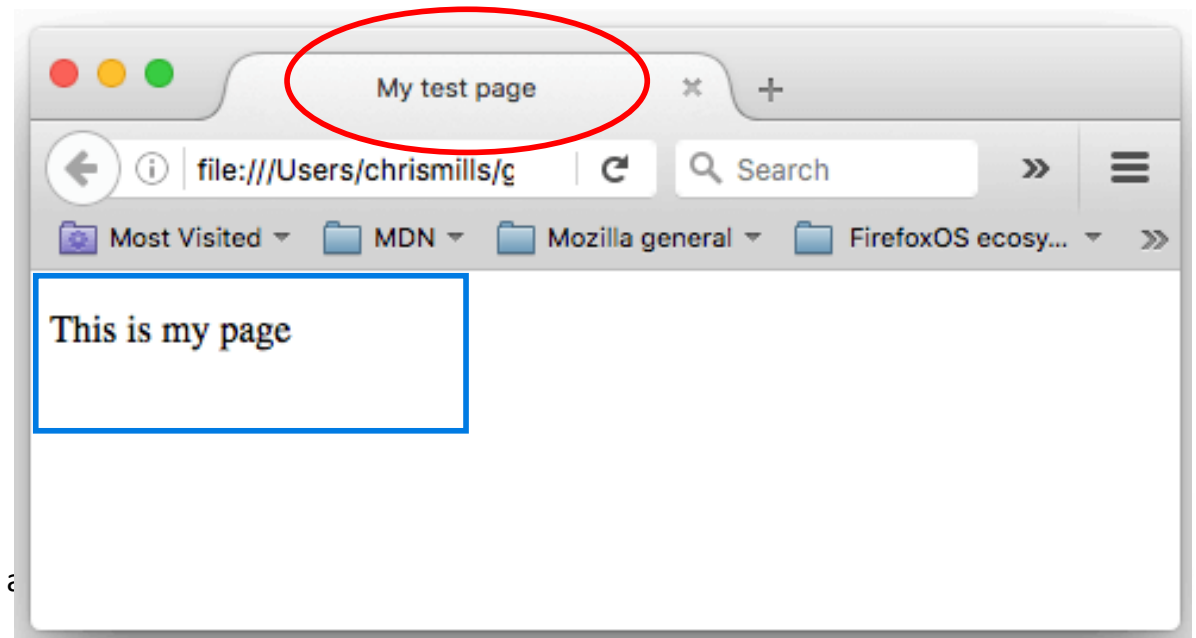
```
<p>This is
```

```
my page</p>
```

```
</body>
```

```
</html>
```

Indentation is irrelevant!  
Multiple spaces, line breaks, tabs  
Are mapped into a single space.



# Structure of an HTML doc

```
<!DOCTYPE html>
```

```
<html>
```

```
<HEAD>
```

```
<meta charset="utf-8">
```

```
<title>My test page</title>
```

```
</head>
```

```
<BOdy>
```

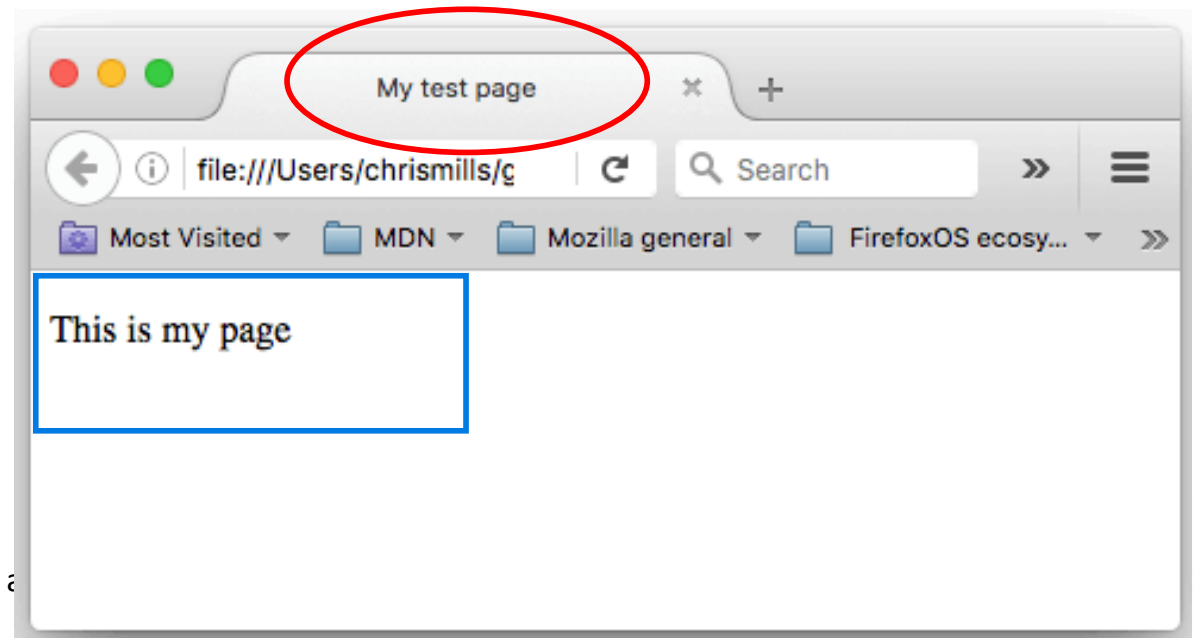
```
<p>This is
```

```
my page</p>
```

```
</body>
```

```
</html>
```

Uppercase and lower case do not matter in tags. (It does in XHTML!)  
Using lowercase is recommended.



# Structure of an HTML doc

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<title>My test page</title>
```

```
</head>
```

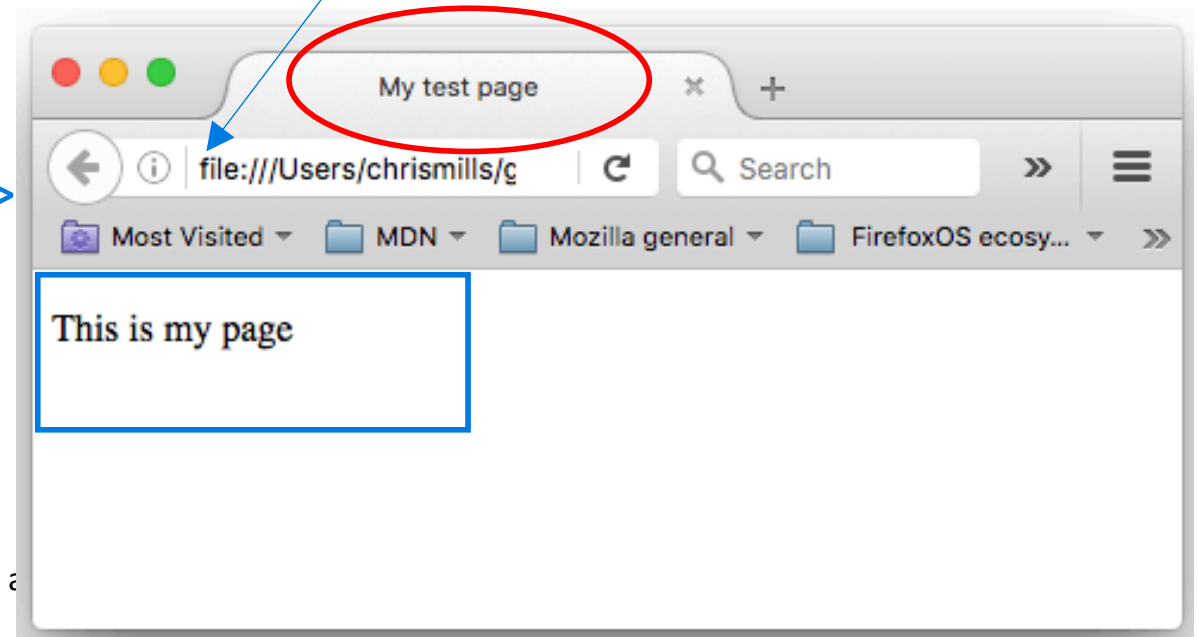
```
<body>
```

```
<p>This is my page</p>
```

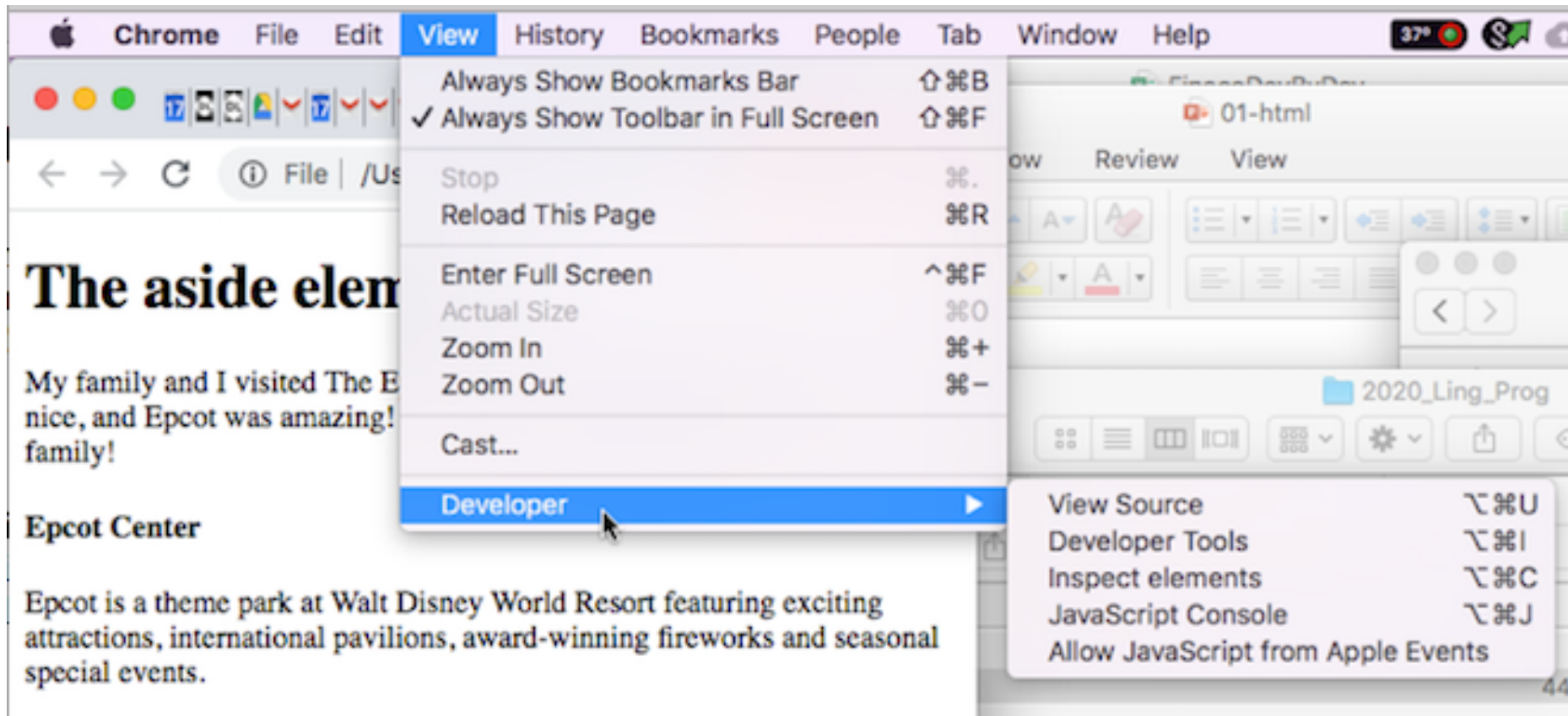
```
</body>
```

```
</html>
```

HTML also works without a server!



# View source



# W3school resources

## Example

<https://www.w3schools.com/html/>

```
<!DOCTYPE html>
<html>
<head>
<title>Page Title</title>
</head>
<body>

<h1>This is a Heading</h1>
<p>This is a paragraph.</p>

</body>
</html>
```

Try it Yourself »



# HTML Editors

HTML editors are called “**WYSIWYG**” (Uizziuaig)

- *What You See Is What You Get!*

Some free, multiplatform HTML Editors:

- Aptana
- Atom
- Bluefish
- Netbeans

(see next slides)



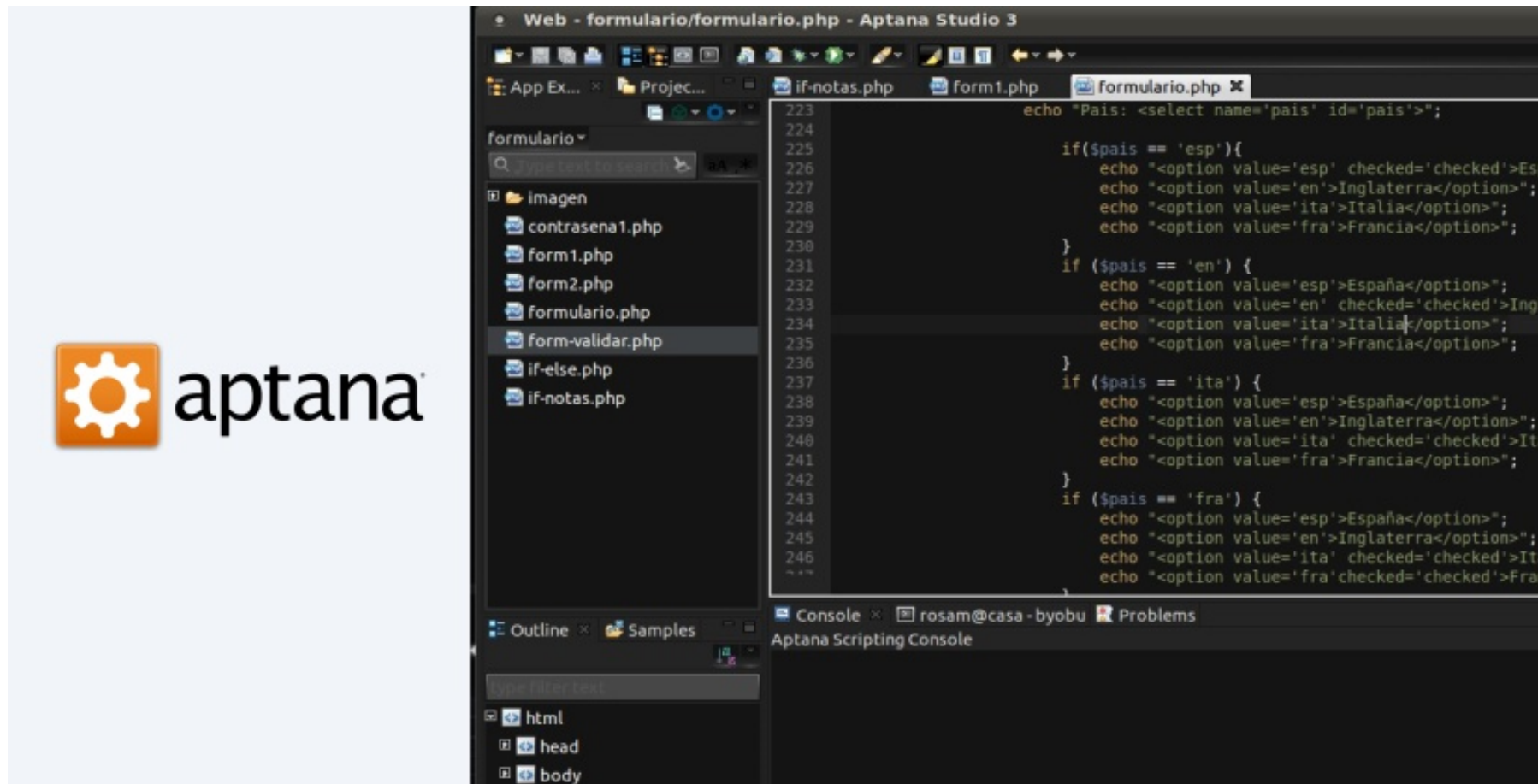
# Multiplatform, free HTML editors - 1



[Brackets](#) - open source text editor.

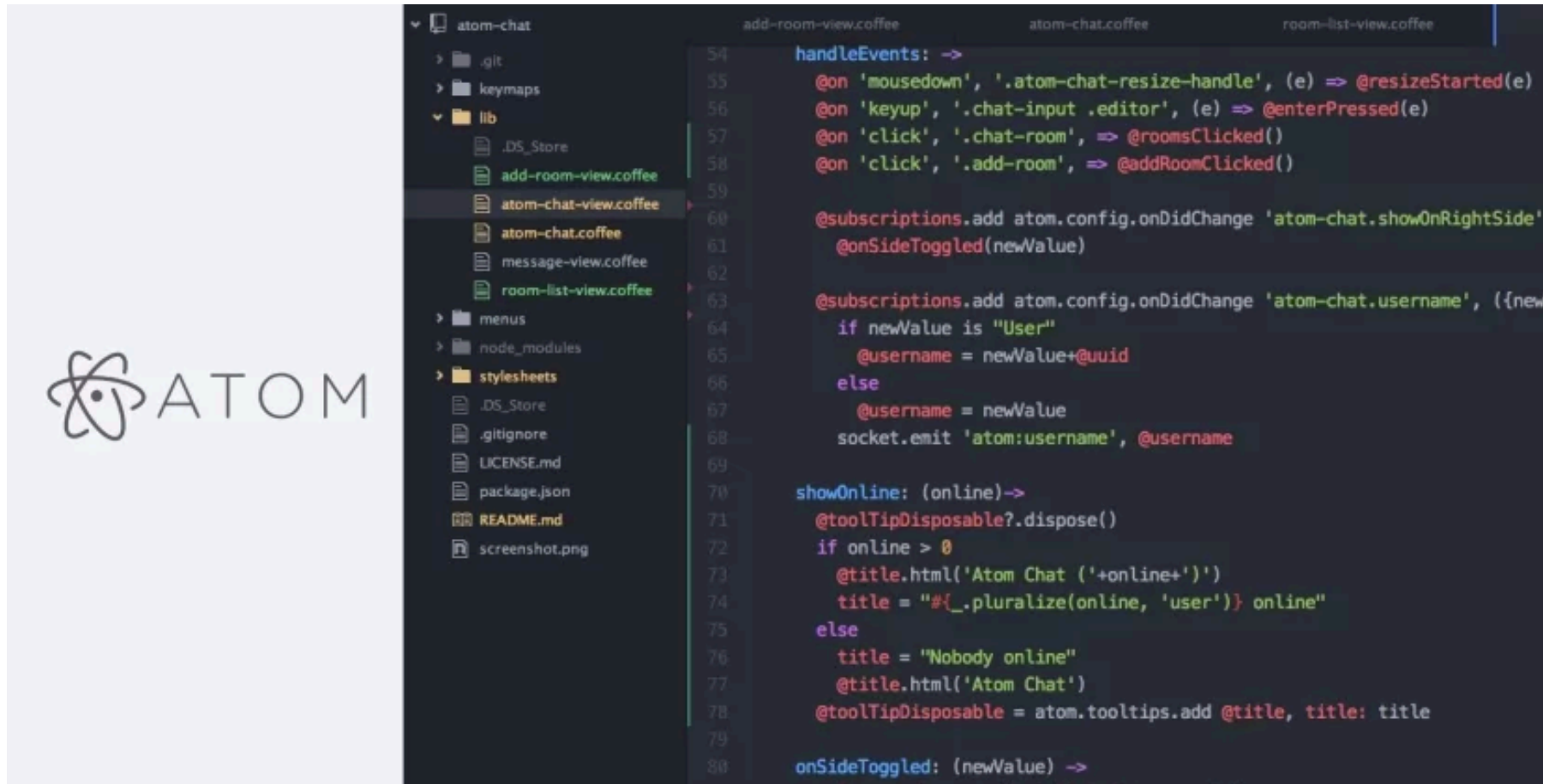


# Multiplatform, free HTML editors - 2



[Aptana Studio 3](#) - open source web application IDE.

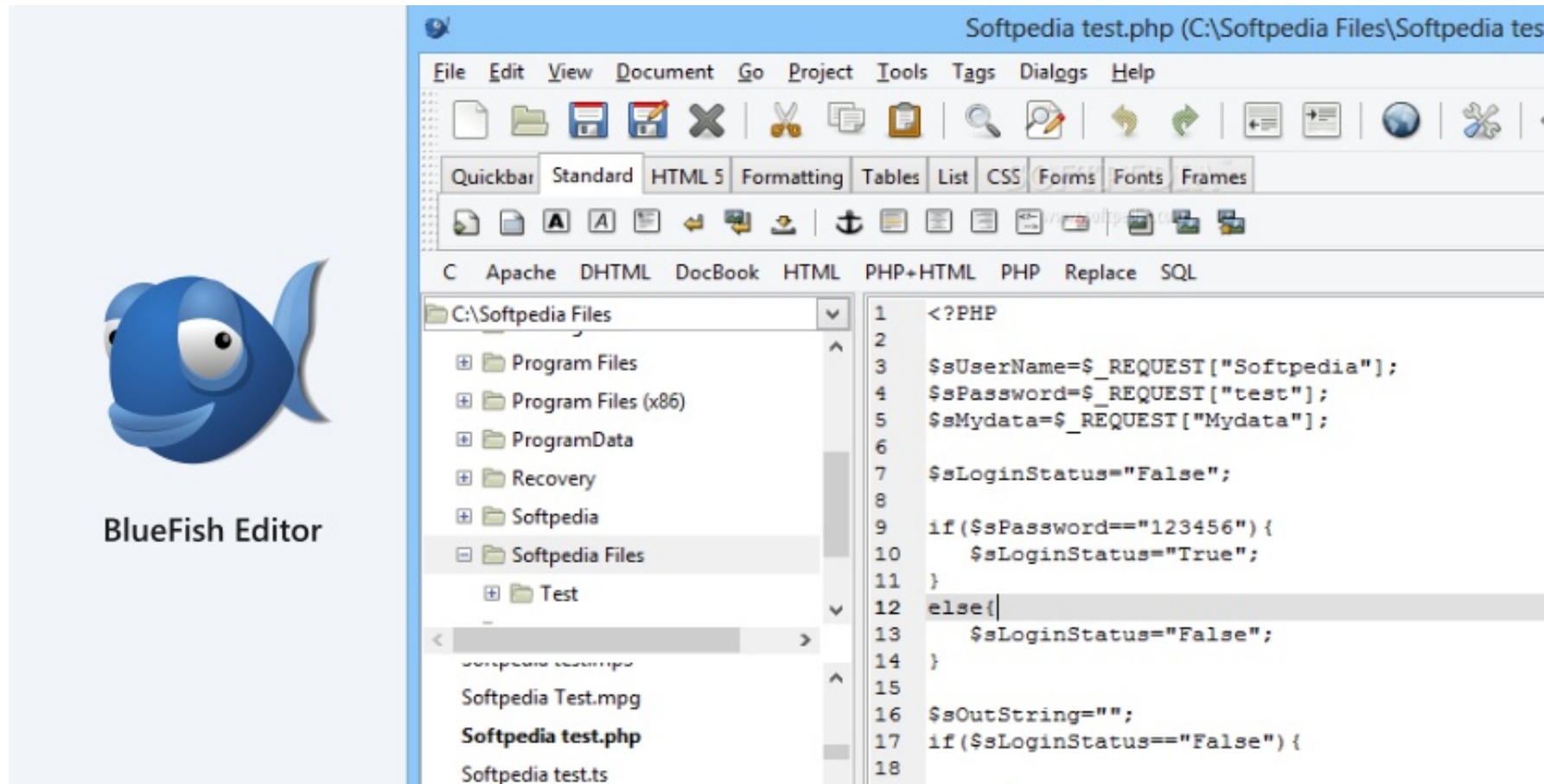
# Multiplatform, free HTML editors - 2



[Atom](https://atom.io/) - open source web application IDE.



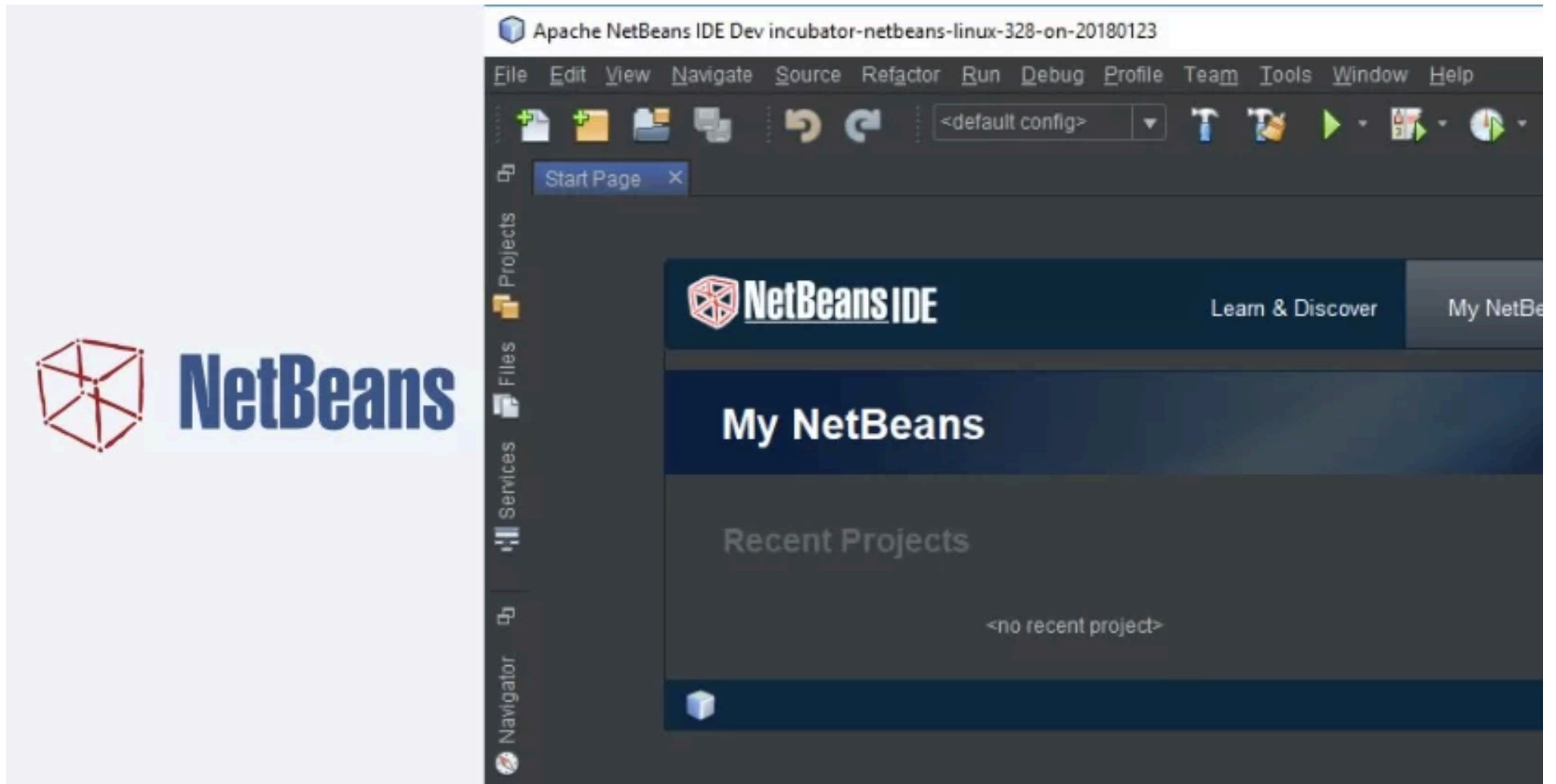
# Multiplatform, free HTML editors – 3



[Kompozer](#) is a free and open source HTML editor with easy to use WYSIWYG editor.



# Multiplatform, free HTML editors - 4



[NetBeans](https://netbeans.org/) contains a free and open source HTML editor.



# HTML Editors pros and cons:

## ■ Pros

- Good for beginners
- You do not have to know much HTML to use editors
- It is easy to create complex tables, image maps, and use advanced functions such as Style Sheets and JavaScript
- Easy for design due to the WYSIWYG interface

## ■ Cons

- They do not always generate clean HTML
  - Add redundant tags
  - Add their own tags
- They do not offer good graphic tools to manipulate images
- It can be hard to fix HTML tags since every time you open a document, code is added.



# HTML Tags

All HTML tags are made up of a tag name, sometimes followed by an optional list of attributes.

**Nothing within the brackets will be displayed by the browser** (unless the HTML is incorrectly written and the browser interprets the tags as part of the content).



# Tags

- A tag is a keyword enclosed by angle brackets ( Example: <B> )
- The opening and closing tags use the same command except the closing tag contains an additional forward slash /
- Example: **<B> Warning </B>**  
causes the word 'Warning' to appear in bold face on a Web page
- Learning HTML essentially means learning a vocabulary of tags.



# Tags – void elements

- There are opening and closing tags for many but not all tags; The affected text is between the two tags.
- See: <https://html.spec.whatwg.org/multipage/syntax.html#void-elements>
- `<BR>` for instance does not strictly require a closing tag – in principle it would be better to use `<BR></BR>` or the **self-closing tag** `<BR/>`

## Is the Ending Slash Optional?

HTML5: the slash is optional.

HTML4: the slash is technically invalid.

However, it's accepted by W3C's HTML validator.

XHTML: The slash is REQUIRED.



# Nested Tags

- Whenever you have HTML tags within other HTML tags, you *should* close the most inner tag first
- Example:

←→  
←→  
<I> <B> The winner is: Dorothea Wierer! </B> Congratulations! </I>  
←→ ***Bold italic*** → ←→ *italic* →

- However, this is legal (NOT in XHTML):

←→  
←→  
<B> The winner is: <I> Dorothea Wierer! </B> Congratulations! </I>  
←→ ***Bold*** → ←→ ***Bold italic*** → ←→ *italic* →

# Attributes

Diagram illustrating HTML attributes in the tag `<body bgcolor="green" text="#000000" link='blue'>`:

- Name**: The attribute name (e.g., `bgcolor`, `text`, `link`).
- Value**: The attribute value (e.g., `"green"`, `"#000000"`, `'blue'`).
- Double quotes**: Indicated by an arrow pointing to the double quotes around the value `"green"`.
- Single quotes**: Indicated by an arrow pointing to the single quotes around the value `'blue'`.

Attributes: Name-value pairs which refine or extend tag's functions

- You can add multiple attributes within a single tag.
- Some attributes only have name (no associated value).
- Values are limited to 1024 characters in length.
- Attributes values are within single or double quotes.



# Basic formatting tags

- Headings **<h1>...<h6>**
- Emphasis **<em>,<strong>,<b>,<i>**
- Subscript, superscript **<sub><sup>**

- Lists **<ol>,<ul>,<li>**

**<ol>**

**<li>...</li>**

**<li>...</li>**

**</ol>**

**OL - Ordered List**

1. First item
2. Second item
3. Third item

**UL - Unodered List**

- First item
- Second item
- Third item

# Basic formatting: blocks

- Line breaks: `<br>`
- Paragraphs and blocks: `<p>`, `<div>`, `<span>`
- **P** and **DIV** have *semantic* difference
  - `<div>` describes a container of data (with graphic separation: has having a line-break before and after it)
  - `<span>` is a logical, in-line grouping without graphic evidence. It's mostly used for a small chunk of HTML inside a line
  - `<p>` describes a paragraph of content (with graphic separation). It cannot contain block-level elements (P, DIV) but can contain SPAN).

*Don't mark up your document based on how it should look - mark it up based on what it is.*

# Advanced formatting

- Quotation `<blockquote>` `<q>`
- Description lists `<dl>``<dt>`
- Details – summary `<detail>` `<summary>` (HTML 5)
- Abbreviation `<abbr>` (HTML 5)
- Address `<address>`
- Computer code `<code>`, `<pre>`, `<var>`, `<kbd>`, `<samp>`
- Time `<time>`

# Images

```

```

## Figure and Figure caption (HTML 5)

```
<figure>  
    
  <figcaption>Cool Ninja Guy</figcaption>  
</figure>
```



Cool Ninja Guy

*How are images indexed by search engines?*

# File paths

<code>&lt;img src="http://machine.domain/images/picture.jpg"&gt;</code>	Absolute Path
<code>&lt;img src="picture.jpg"&gt;</code>	picture.jpg is located in the same server, same folder as the current page
<code>&lt;img src="images/picture.jpg"&gt;</code>	picture.jpg is located in the images folder in the current folder
<code>&lt;img src="/images/picture.jpg"&gt;</code>	picture.jpg is located in the images folder at the root of the current web
<code>&lt;img src="../picture.jpg"&gt;</code>	picture.jpg is located in the folder one level up from the current folder

# Hyperlinks

`<a href="URL">testo del link</a>`

- Also URLs with other protocols:

`<a href="mailto:marco.ronchetti@unitn.it">write to me!</a>`

`<base href="https://www.mydomain.com/" target="_blank">`

# Hyperlinks – inner target

```
<!DOCTYPE html>
```

```
<html>
```

```
<body>
```

```
<h2>The inner target</h2>
```

```
<a href="#pippo" >GO TO PIPPO!</a>
```

```
<p>one of many lines...</p>
```

```
...
```

```
<p>one of many lines...</p>
```

```
<a name="pippo"/>
```

```
<H1>PIPPO IS HERE!</H1>
```

```
<p>one of many lines...</p>
```

```
...
```

```
<p>one of many lines...</p>
```

```
</body>
```

```
48  
</html>
```

# Iframes - comments

- Comments `<!-- This is a comment -->`

- Iframes

```
<iframe src="demo_iframe.htm"></iframe>
```

```
<iframe height="300px" width="100%" src="demo_iframe.htm"  
name="iframe_a"></iframe>
```

```
<p><a href="https://www.w3schools.com"  
target="iframe_a">W3Schools.com</a></p>
```

# Character Encoding

Describes the mapping between chars and integer numbers

- **ASCII** was the first **character encoding standard** (also called character set). 128 different alphanumeric characters: {numbers (0-9), English letters (A-Z), some special characters e.g. ! \$ + - ( ) @ < >, some non printable ones like ESC}.
- **ISO-8859-1**: default character set for HTML 4. This character set supported 256 different character codes.
- **ANSI** (Windows-1252): the original Windows character set. Added 32 extra characters to ISO-8859-1.
- **UTF-8** (Unicode) was supported by HTML4, is the default character set for HTML5.

# Entities

Result	Description	Entity Name	Entity Number
	non-breaking space	&nbsp;	&#160;
<	less than	&lt;	&#60;
>	greater than	&gt;	&#62;
&	ampersand	&amp;	&#38;
"	double quotation mark	&quot;	&#34;
'	single quotation mark (apostrophe)	&apos;	&#39;
¢	cent	&cent;	&#162;
£	pound	&pound;	&#163;
¥	yen	&yen;	&#165;
€	euro	&euro;	&#8364;
©	copyright	&copy;	&#169;
®	registered trademark	&reg;	&#174;

# URL Encoding

Character	From Windows-1252	From UTF-8
space	%20	%20
!	%21	%21
"	%22	%22
#	%23	%23
\$	%24	%24
%	%25	%25
&	%26	%26
'	%27	%27
(	%28	%28
)	%29	%29

- See also [https://www.w3schools.com/tags/ref\\_urlencode.asp](https://www.w3schools.com/tags/ref_urlencode.asp)

# Metadata (head)

```
<meta charset="UTF-8">
```

```
<meta name="description" content="Free Web tutorials">
```

```
<meta name="keywords" content="HTML,CSS,XML,JavaScript">
```

```
<meta name="author" content="John Doe">
```

```
<meta http-equiv="refresh" content="30">
```

```
<meta charset="utf-8"> Charset declaration
```

- Link CSS/JS

# Metadata (head)

```
<meta name="viewport" content="width=device-width, initial-scale=1.0">
```

With viewport, on phone



Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Nam liber tempor cum soluta nobis

Without viewport, on phone



# What I expect you to know...

Ignore (for now) references to “styles”

HTML Tutorial	<del>HTML CSS</del>	HTML Computercode
HTML HOME	HTML Links	HTML Entities
HTML Introduction	HTML Images	HTML Symbols
HTML Editors	HTML Tables	HTML Charset
HTML Basic	HTML Lists	HTML URL Encode
<b>HTML Elements</b>	HTML Blocks	<del>HTML XHTML</del>
HTML Attributes	<del>HTML Classes</del>	<del>HTML Forms</del>
HTML Headings	<del>HTML Id</del>	
HTML Paragraphs	HTML Iframes	
<del>HTML Styles</del>	<del>HTML JavaScript</del>	
HTML Formatting	HTML File Paths	
HTML Quotations	HTML Head	
HTML Comments	<del>HTML Layout</del>	
HTML Colors	<del>HTML Responsive</del>	

# Tables

`<table>` defines the tabella

`<th>` or `<tr>` define the rows

`<td>` defines every single cell

`<caption>` defines the caption (didascalia)

The attributes

- `colspan="n"`
- `rowspan="n"`

define the cell span over columns and rows.

# Tables

## Example

`<h2>Cell that spans two columns</h2>`

```
<table style="width:100%">
  <tr>
    <th>Name</th>
    <th colspan="2">Telephone</th>
  </tr>
  <tr>
    <td>Bill Gates</td>
    <td>555 77854</td>
    <td>555 77855</td>
  </tr>
  <tr>
    <td>Donald Duck</td>
    <td>945 2626</td>
    <td>0471 23706</td>
  </tr>
</table>
```

### Cell that spans two columns

Name	Telephone	
Bill Gates	555 77854	555 77855
Donald Duck	945 2626	0471 23706

# Homework -1

Create a webpage that looks like this: 

The image can be retrieved from  
[https://www.mrwebmaster.it/img/cope\\_tax/2.jpg](https://www.mrwebmaster.it/img/cope_tax/2.jpg)

The index items should point to the corresponding sections below.

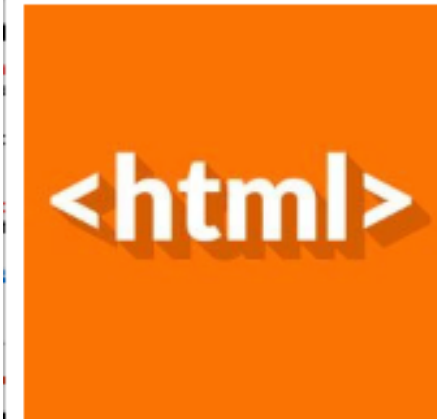
The final link should connect to  
<https://piazza.com/unitn.it/spring2020/programmazioneweb/home>.

## My First Website

### Indice

1. [Images](#)
2. [Formatting](#)
3. [Lists](#)
4. [Links](#)

### Images



### Formatting

This is an example of **bold**, *italic*, ***bold italic***,

### Lists

List of tags used in the body of this page

- `<h1>`, `<h2>`
- `<ol>`, `<ul>`, `<li>`
- `<a>`
- `<b>`, `<i>`
- dragonfruit

### Links

[The web site of this course on Piazza](#)

# Homework -2

Change the webpage so that it takes the following aspect:

## My First Website

### Indice

1. [Images](#)
2. [Formatting](#)
3. [Lists](#)
4. [Links](#)

### Images



### Formatting

This is an example of **bold**, *italic*, ***bold italic***,

### Lists

List of tags used in the body of this page

- `<h1>`, `<h2>`
- `<ol>`, `<ul>`, `<li>`
- `<a>`
- `<b>`, `<i>`
- dragonfruit

### Links

[The web site of this course on Piazza](#)

# Homework -3

Create a webpage that looks like this:

## Cells that spans over rows and columns

1,1	2,1	3,2	4,1	5,1	6,1	7,1	8,1	9,1	10,1
1,2	2,2		4,2	5,2	6,2	7,2	8,2	9,2	10,2
			4,3	5,3	6,3	7,3	8,3	9,3	10,3
	2,4		4,4			7,4	8,4	9,4	10,4
	2,5					7,5	8,5	9,5	10,5
	2,6					7,6	8,6	9,6	10,6
	2,7		4,7			7,7			
	2,8								
	2,9								
	2,10								

Suggestion: bgcolor="red"