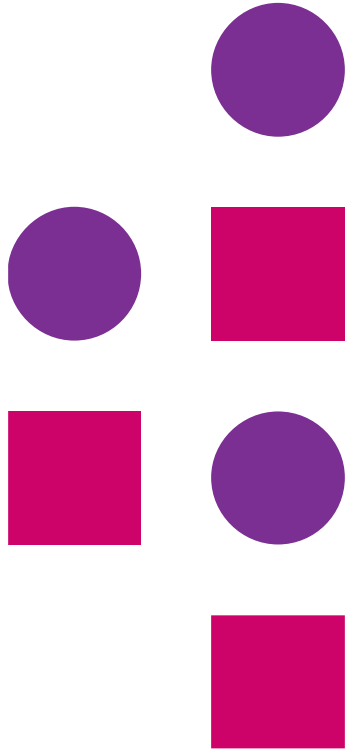
WSA | Forward looking IT finishing school

Entities, Symbols, Charset

Hypertext Markup Language 5 (HTML5)

Table of Content

- Entities
- Symbols
- Charset



Entities

(Hypertext Markup Language 5)

Entities

- Some characters are reserved in HTML
- For example less than(<) or greater than(>) characters
- Browser might mix such characters with tags which would lead to unexpected problems
- HTML reserved characters, if to be displayed to user, must be replaced with character entities
- Characters that are not present in keyboard can also be replaced by entities

Character Entity

Syntax :
&entity-name; // Name
or
&#entity-number; // Number

HTML – Character Entity Reference

Symbol	Entity Name	Entity Number	Description
&	&amp;	&#38;	Ampersand
<	&lt;	&#60;	Less than symbol
>	&gt;	&#62;	Greater than symbol
“	&quot;	&#34;	Double quotation
©	&copy;	&#169;	Copy Right
®	&reg;	&#174;	Registered Trademark
™	&trade;	&#8482;	Trademark
			Non-breaking space
‘	&apos;	&#39;	Apostrophe (Single Quote)

Entities

- Advantages
  - An entity name is easy to remember
- Disadvantages
  - Browser may not support all entity names (but support for number is good)

HTML Symbols

- Many Mathematical, technical and currency symbols, are not present on a normal keyboard
- To add such symbols to an HTML page you can use an HTML Entity name
- If no entity name exists, we can use an entity number, a decimal, or hexadecimal reference

HTML Symbols

- **Example**
- European currency : €
- European currency : €
- Indian currency:&inr
- Indian currency : ₹
- **Output**
- European currency : €
- European currency : €
- Indian currency : ₹

HTML – Classwork

- Write a web page to display following text

This is a "paragraph"

10 > 5

6 < 13

Intel® and Microsoft® are registered trademarks of Intel and Microsoft.

©Webstack Academy (P) Ltd.

Pepsi™ is trademarks of PepsiCo India

Charset

(Hypertext Markup Language 5)

Character Encoding

- A character encoding is method of converting bytes into characters
- Popular encoding techniques are ASCII, ANSI and ISO-8859-1
- ASCII was the first character encoding standard which defined 127 different alphanumeric characters
- ANSI was the original Windows character set, with support for 256 different character code
- ISO-8859-1 was the default character set for HTML4
- ANSI & ISO-8859-1 were limited, so default character encoding was changed to UTF-8 in HTML5

The Charset Attribute

- To display an HTML page correctly, a web browser must know the character set used in the page
- Charset is specified in the <meta> tag
- <meta charset=”UTF-8”>

URL Encoding

(Hypertext Markup Language 5)

URL (Uniform Resource Locator)

- A Uniform Resource Locator (URL) is used to address a document on the web
- A web address follows these rules
  - “scheme://prefix.domain:port/path/filename”

URL

- scheme – defines the type of Internet service (http or https)
- prefix – defines a domain prefix (www)
- domain - defines the Internet domain
- port - defines the port number at the host (default for http is 80)
- path - defines a path at the server
- filename - defines the name of a document or resource

URL Encoding

- URL encoding converts characters into a format that can be transmitted over Internet
- Any attribute with a value that is a URL must be URL-encoded
- Example
  - <a href>
  - <img src>
  - <iframe src>

URL Encoding

- URLs can only be sent over the internet using the ASCII character-set
- URL encoding replaces unsafe ASCII characters with a “%” followed by two hexadecimal digits
- URLs cannot contain spaces. URL encoding normally replaces a space with a plus(+) sign or with %20

WSA | Forward looking IT finishing school

Web Stack Academy (P) Ltd

#83, Farah Towers,
1st floor,MG Road,
Bangalore – 560001

M: +91-80-4128 9576
T: +91-98862 69112