



Character encoding
Character encoding

Each character is assigned a particular number called a **code point**. These code points are stored in computer memory in the form of bytes (a unit of data in computer memory). In technical terms, we say the character is **encoded** using one or more bytes.

Basically, all the characters are stored in computer language and a **character encoding** is the awesome dictionary that is going to help us decode this computer language into something we can understand. In technical terms, it is what is used as a reference to map code points into bytes to store in computer memory; then when you use a character in your HTML, the bytes are then read back into code points using the character encoding as a reference.

- *ASCII*: contains letters, characters and a limited set of symbols and punctuation for the English language
- *Windows-1252 (Latin1)*: Windows character set that supports 256 different code points
- *ISO-8859-6*: contains letters and symbols based on the Arabic script

- *Unicode*: contains characters for most living languages and scripts in the world

When you code in HTML, you must specify the encoding you wish your page to use. Providing no encoding or the wrong one is pretty much like providing the wrong dictionary to decode. It can display your text incorrectly or cause your data to not be read correctly by a search engine. A character encoding declaration in your HTML is also important to process unfamiliar characters entered in forms by users, URLs generated by scripts, etc.

You should always use the Unicode character encoding UTF-8 for your Web pages, and avoid 'legacy' encodings such as ASCII, Windows-1252 and ISO-8859-6 mentioned above. Do not use the UTF-16 Unicode encoding either.

It is important to note that it is not enough to simply declare your encoding at the top of the web page. **You have to ensure that your editor saves the file in UTF-8** also. Most editors will do that these days, but you should check.

Read an [Introduction to character sets and encodings here](#).