Unicode

Unicode is a universal character set, ie. a standard that defines, in one place, all the characters needed for writing the majority of living languages in use on computers. It aims to be, and to a large extent already is, a superset of all other character sets that have been encoded.

Text in a computer or on the Web is composed of characters.*Characters* represent letters of the alphabet, punctuation, or other symbols.

As a content author or developer, you should nowadays always[choose the UTF-8 character encoding](#) for your content or data. This Unicode encoding is a good choice because you can use a single encoding to handle pretty much any character you are likely to meet. This greatly simplifies things.