

Non-ASCII characters in ARCHER 3.2

The master-text is coded in UTF-8, and the text version in Latin-1, so most special characters such as accented and Greek letters and <£, ½>, etc. are stored and displayed directly. The master version of the corpus currently contains the following five HTML entities (with frequencies):

character	master files	CQPweb display	text files	N
apostrophe	'	'	'	46822
straight double quotation mark	"	"	"	31605
ampersand	&	&	&	7663
less than	<	< [blank in list of hits]	<	144
greater than	>	>	>	123

(corrected 4-Jan-14)