**UTF-8** is a [variable width](https://en.wikipedia.org/wiki/Variable-width_encoding) [character encoding](https://en.wikipedia.org/wiki/Character_encoding) capable of encoding all 1,112,064[[1]](https://en.wikipedia.org/wiki/UTF-8#cite_note-1) valid [code points](https://en.wikipedia.org/wiki/Code_point) in [Unicode](https://en.wikipedia.org/wiki/Unicode) using one to four 8-bit [bytes](https://en.wikipedia.org/wiki/Byte).[[2]](https://en.wikipedia.org/wiki/UTF-8#cite_note-2) The encoding is defined by the Unicode standard, and was originally designed by [Ken Thompson](https://en.wikipedia.org/wiki/Ken_Thompson) and [Rob Pike](https://en.wikipedia.org/wiki/Rob_Pike).[[3]](https://en.wikipedia.org/wiki/UTF-8#cite_note-:0-3)[[4]](https://en.wikipedia.org/wiki/UTF-8#cite_note-4) The name is derived from *Unicode* (or *Universal Coded Character Set*) *Transformation Format – 8-bit*.[[5]](https://en.wikipedia.org/wiki/UTF-8#cite_note-5)

It was designed for [backward compatibility](https://en.wikipedia.org/wiki/Backward_compatibility) with [ASCII](https://en.wikipedia.org/wiki/ASCII). Code points with lower numerical values, which tend to occur more frequently, are encoded using fewer bytes. The first 128 characters of Unicode, which correspond one-to-one with ASCII, are encoded using a single octet with the same binary value as ASCII, so that valid ASCII text is valid UTF-8-encoded Unicode as well. Since ASCII bytes do not occur when encoding non-ASCII code points into UTF-8, UTF-8 is safe to use within most programming and document languages that interpret certain ASCII characters in a special way, such as "[/](https://en.wikipedia.org/wiki/Slash_(punctuation))" in filenames, "[\](https://en.wikipedia.org/wiki/Backslash)" in [escape sequences](https://en.wikipedia.org/wiki/String_literal#Escape_sequences), and "%" in [printf](https://en.wikipedia.org/wiki/Printf).