

MORETF8

DAN BRUMLEVE

1. OVERVIEW

moretf8 an extension of utf8 that supports embedded binary data.

The bytes 0xFF and 0xFE do not appear in any valid utf8 byte sequence. An atomic utf8 byte sequence is the utf8 encoding of a single Unicode character, and a utf8 byte sequence can be considered as a sequence of atomic utf8 byte sequences. An atomic moretf8 byte sequence is either an atomic utf8 byte sequence, or it is 0xFF followed by some utf8 character followed by a byte sequence whose length is that character's codepoint.