

Exercises for chapter: Character encoding

1. Show that a UTF-8 parser will not miss more than two characters if a byte becomes damaged (any number of bits arbitrarily changed).
2. What does an `ALT` key do?
3. What is EBCDIC? What is the basic idea?
4. Find the Unicode definition. Can you find an example of a character that has two functions, but is not defined as two characters? Find two characters that are defined separately for compatibility, but that are defined equivalent.
5. ISO 8859 has the ‘non-breaking space’ at position `A0`. How does `TEX` handle the `nbsp`? How do `TEX`, HTML, Latin-1, MS Word, et cetera handle multiple spaces? Discuss the pros and cons.