

# Introduction to Text

In computer systems, text is represented in files as a sequence of characters, each of which corresponds to a specific number, known as a character code. These character codes are then stored in the file as binary data.

## 1.1 Newlines and Carriage Returns

Two of the character codes that have special meanings are the newline (often represented as `'\n'`) and the carriage return (often represented as `'\r'`).

The newline character signifies the end of a line of text and the beginning of a new one. If you parse a text file with a Python script, it will see your 'enter' key press as a new line symbol. The carriage return character moves the cursor to the beginning of the line. The use of these characters can vary between operating systems. Unix-based systems (like Linux and MacOS) use the newline character to indicate the end of a line, while Windows systems use a combination of a carriage return and a newline (`'\r\n'`).

## 1.2 ASCII

The American Standard Code for Information Interchange (ASCII) is one of the earliest character encodings. It uses 7 bits to represent each character, allowing it to define up to  $2^7 = 128$  different characters. These include the English alphabet (in both lower and upper cases), digits, punctuation symbols, control characters (like newline and carriage return), and some other symbols.

## 1.3 UTF-8

UTF-8 (8-bit Unicode Transformation Format) is a variable-width character encoding that can represent every character in the Unicode standard, yet remains backward-compatible with ASCII. For the ASCII range (0-127), UTF-8 is identical to ASCII. However, it can use additional bytes (up to 4 bytes in total) to represent characters that are not included in ASCII, such as characters from other languages, emojis, and many other symbols. This has made UTF-8 a widely used encoding in many modern systems.

*This is a draft chapter from the Kontinua Project. Please see our website (<https://kontinua.org/>) for more details.*

# Answers to Exercises





---

# INDEX

ASCII, [1](#)

carriage return, [1](#)

newline, [1](#)

text, [1](#)

UTF-8, [1](#)