Character Encoding

Chapter No.	21
Status	Completed

▼ A byte is the number of bits used to encode a single character of text in a computer

- A byte is usually comprised of eight bits

- A character set is the complete set of characters that a particular computer uses

▼ The American Standard Code for Information Interchange (ASCII) is a character encoding standard for electronic communication.

- ASCII codes represent text in computers, telecommunications equipment, and other devices

▼ Unicode is an information technology standard for the consistent encoding, representation, and handling of text expressed in most of the world's writing systems

- The ultimate aim of Unicode is to be able to present any possible text in any written language in code form
- This has been extended to include a number of other symbols used in technical situations, as well as emoji
- Unicode is designed so that once a code has been determined, it never changes
- A character code is referred to as a code point

▼ ASCII & Unicode in Python

▼ chr( )

- Returns the character that represents the specified ASCII code or Unicode code point

▼ ord( )

- Returns the number that represents the ASCII code or Unicode code point of the specified character
- In Python, Unicode code points are represented using the escape sequence "\u", followed by the hexadecimal numbers of the code point