Topic: Text and Character Sets

Representation of Text and Character Sets (ASCII and Unicode)

Computers process and store all data, including text, in binary form. To represent text, character sets such as ASCII (American Standard Code for Information Interchange) and Unicode are used to map individual characters to unique binary codes.

Text Conversion to Binary:

Text is converted to binary using predefined character sets that assign a specific binary code to each character. This allows computers to process, store, and transmit text as binary data.

ASCII:

ASCII is a widely used character encoding that represents 128 characters, including uppercase and lowercase English letters, digits, punctuation marks, and control characters, using 7 bits per character. It is limited in its ability to represent characters from other languages or additional symbols.

Unicode:

Unicode is a more extensive character encoding that supports a wide range of characters and symbols, including those from different languages, mathematical notations, and emojis. It overcomes the limitations of ASCII by providing unique codes for over a million characters. Unicode requires more bits per character compared to ASCII, that is 16bits per character.