

Text Representation

All **data** stored in a computer is stored as binary numbers.

This means that writing (text) has to be stored as numbers. A different number **represents** each key on a keyboard.

Two ways to do this are **ASCII code** and **Unicode**.

Text Representation

Text is made up of individual **characters**
(letters, numbers, spaces etc...)

The number used to store a specific character is called the **character code**

The set of character codes you have available to use is called a **character set**

ASCII code

ASCII code uses numbers to encode text

- first used 1963 for teleprinter machines
- uses **7 bit numbers** (0-127 = 128 possible)
- standard American English keyboard characters only - this is the **character set**
- includes keyboard commands - space, new line, new paragraph, quit etc...
- simple and quick to transmit

ASCII code



ASCII code



ASCII code

Each **character code** in ASCII is a 7 bit number (0-127)

1. How many bits are needed to store the word “binary”?
2. How many bits are needed to store the word “Woolloomooloo”?

ASCII code

1. How many bits are needed to store the word “binary”?

6 character of 7 bits each = $6 \times 7 = 42$ bits

2. How many bits are needed to store the word “Woolloomooloo”?

12 characters = $12 \times 7 = 94$ bits (repeated characters still need to be stored)

ASCII code

073 032 108 105

107 101 032 115

079 085 112 033

ASCII code

Letters are grouped - 65 to 90 are capital letters; 97 to 122 are lower case letters

IMPORTANT:

- “A” is 065
- “a” is 097

These are **not the same thing** in a computer

“A” != “a” - this is important in Python coding

ASCII code

Letters are grouped - 65 to 90 are capital letters; 97 to 122 are lower case letters

Q. If the character code 077 represents the letter M:

- a) Write down the character code which will represent the letter R
- b) Which letter is represented by the character code 074

ASCII code

Q. If the character code 077 represents the letter M:

- a) Write down the character code which will represent the letter R = 082 (077 + 5 as R is five letters on from M)
- b) Which letter is represented by the character code 074 = J (074 is three places before 077; J is three places before M in the alphabet)

ASCII code

Character code 032 is a space. A space is a thing (in programming, it is a string of length 1)

Character codes 0 – 31 and 127 are **non-printable**

These were required to allow teleprinter machines to manage messages properly.

A lot of these codes also appear on a keyboard - for example delete, backspace and return