File handling →

Agenda → How Python stores, write/Read any file → data

→ DE/ Heavy dataset Companies → ⁂

File handling →

⇒ functions, Classes, objects → where/ How are they stored.

Memory

Primary (RAM) (ROM)

→ Volatile / Temporary Storage

→ [working memory]

Secondary (HD, SDD, CD, DVD, ...) USB drive

→ persistant/ permanent

[what you save]

⇒ Does machine understand english??

↳ No → 0, 1

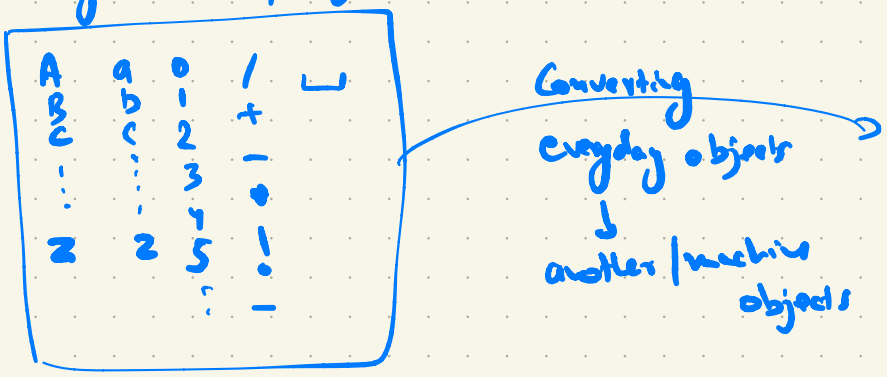
Store [abc] → into a memory

text How??

⇒ Encoding →

↓

Creating a mapping in a predefined format.



A → 010101

B → 010110

C → 010111

D → 011001

0 → ——

→ ——

→ ——

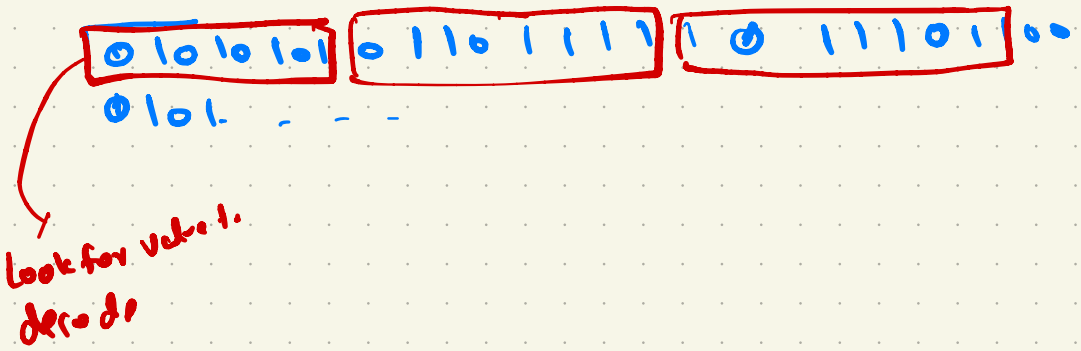
A → 010101
B → 010110
C → 010111
D → 011001

ABC → 010101 010110 010111

By default → machine don't know where to start / end for each character.

→ Fix the number of 0/1 in the encoding.

every character → alpha, Numeric, Symbol → fixed length



Q How many values can be stored in 8 0/1s:

□□□□□□□□ $2^8 = 256$ distinct characters

1 2 ... ——

⇒ 8 bit encoding → 256 characters

ASCII

⇒ 8 bits → bytes →

Every ascII element takes 1 byte spare

original ascII → 128 char + extended

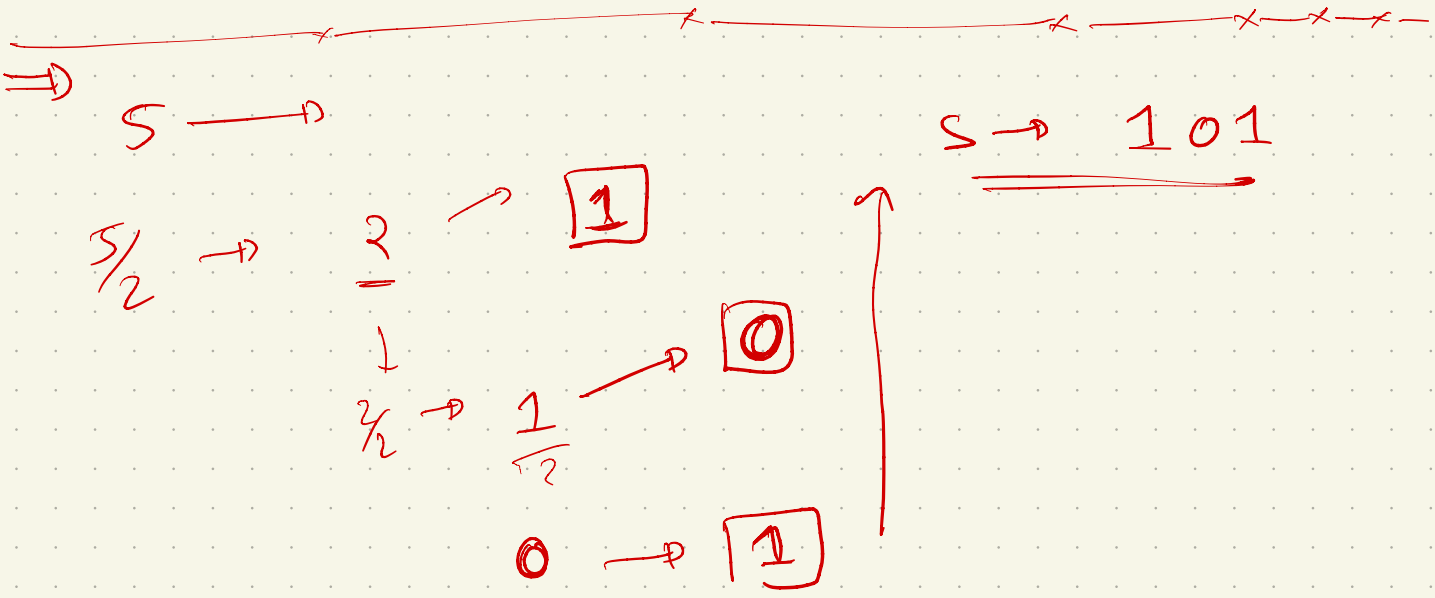
extended ascII → 256 char

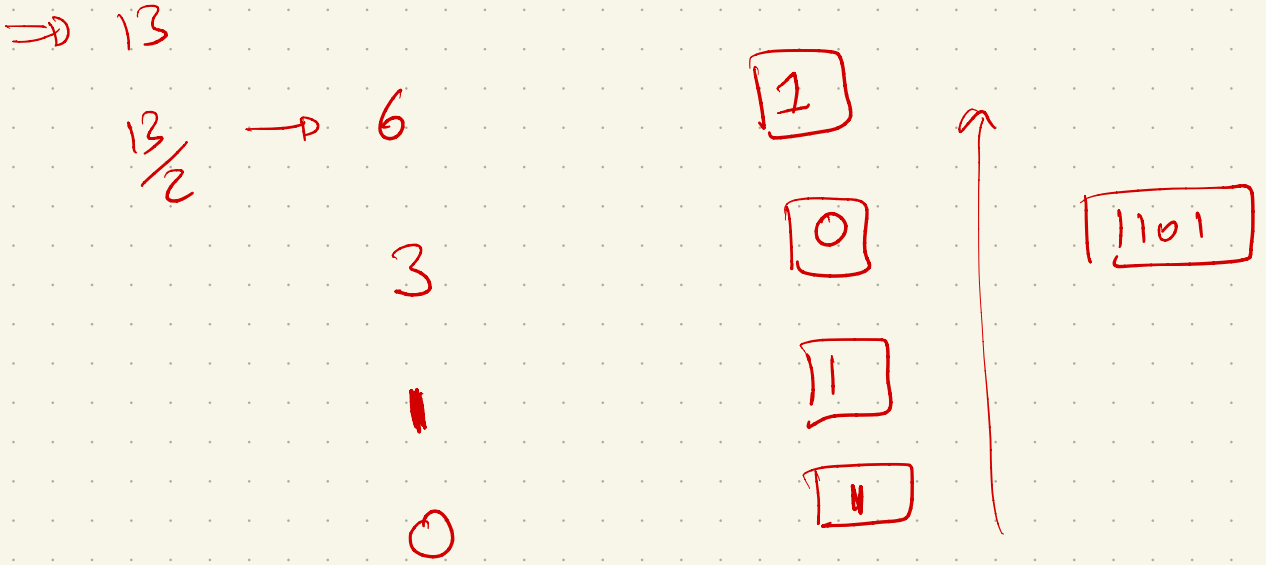
→ UTF-8
UNICODE
ANSI

Given Character → binary Code

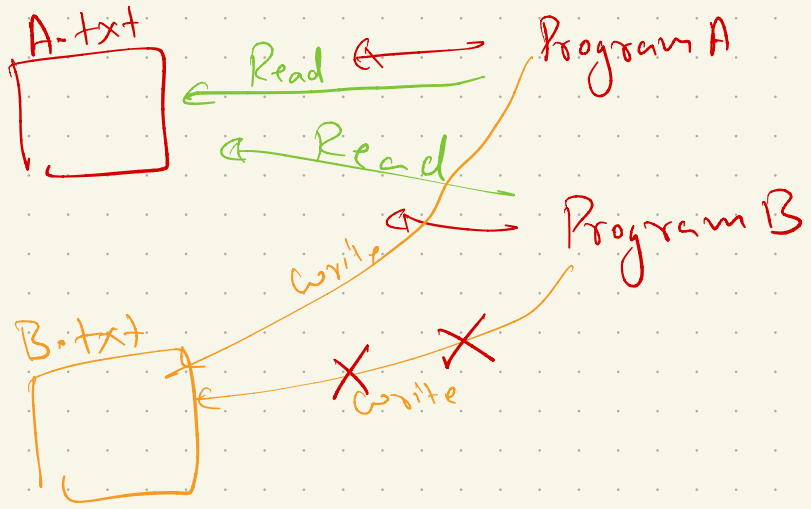
① Initial characters of ascII aswell as UTF are same.

UTF-8 encoding





⟹ When to allow read / write.

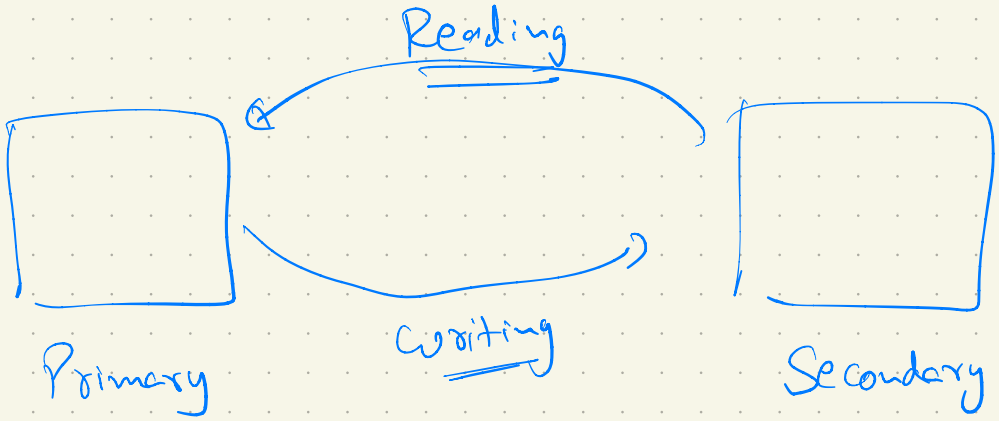


Read → Simultaneously

write → only 1 at a time.

⟹ Should we allow 100 programs to read the file?? [Most programs/ machine have a failsafe]

Read / write →



⇒ File access Modes →

1) Read only (r) → Open file for reading.

2) Write only (w) → Only write but don't read the data.

3) Read and write (r+) → read and also written.
[If file doesn't exist → Error]

4) write and read (w+) → write and read from file
(data is overwritten if exists) [If file doesn't exist → Create the file]

5) Append (a) → any new data is added at the end of the file

⇒ Why closing the file is required →

(1) Improve performance → freeing up the memory.

(2) Prevent Corruption →
↳ Most languages/ programs only write/save the data to disk when you close the file.

⇒ Everytime we are doing write →
data is overwritten !.

⇒ Readline → 1 by 1 → slow process

⇒ All read|write|append → use default encoding.

⇒ Image
Video
Audios
→ are not utf-8 encoded.
↳ Jpg, png, mp4, ... → encoder

⇒ write file → overwrite ?
append file → add at end !.

① How to add something in middle of the file

② File.read, readlines, - - →
How to go back and read again ??

⇒ Closing file everytime is a hassle.

⇒ File → Abc.txt

A B C D E F G H

I J K L M N O P

→ write → replace from the beginning

↳ only till the length of input

https://colab.research.google.com/drive/1wogxUFaTS9fBctFqh6zxZN2VNzUJiqQp?usp=sharing