Python Programming
Files

Mostafa S. Ibrahim
Teaching, Training and Coaching since more than a decade!

Artificial Intelligence & Computer Vision Researcher
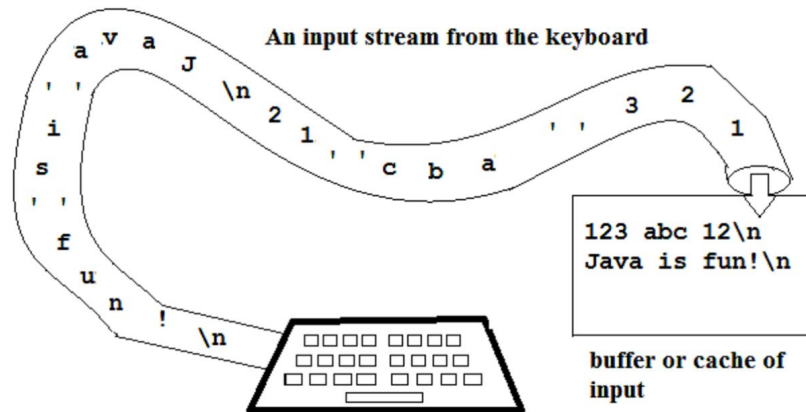*PhD* from Simon Fraser University - Canada
Bachelor / Msc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)

From Console to Files

- So far we read & write to **console** (the black screen / IDL screen)
  - IO stream from keyboard using input() and print()
- In practice, we want to **save data** in disk **not memory**
- We should stream (read/write) to files
- Files can be used to store the data.
  - We use Files all the time
  - Data.txt, mostafa.jpg, python.exe
- Databases are an advanced mechanism to store the data
  - rdbms, nosql

An input stream from the keyboard

a v a J
i
s
f
u
n !
\n
2 1
c b a
3 2
1

123 abc 12\n
Java is fun!\n
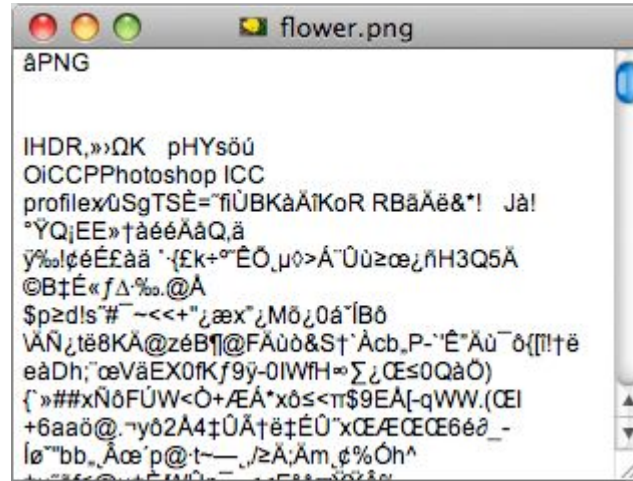
buffer or cache of input

Img src

File Structure & Types

- File is typically **3 parts**
  - **Header**: metadata such as file name and its type
  - **Data**: for the content (encoded binary or text)
  - **End of file (EOF)**: A special character to mark the file end

Binary file vs Text file

- Storing data as sequence of bytes rather than text (a..z, 0, 1, 2…#%...)
  - Not readable in editors
  - You need to read the content according to the way it was wrote.

Binary file vs Text file

- Text files
    - Limited to natural text
    - If some letter is corrupted, file is still opening, You even spot with eyes
    - Buy file size is bigger, and vary in size based on platform
- Binary files
    - Allow text and binary data
    - A minor corruption = probably failure to get back data
    - File size is smaller

File name and path

- Each file has a name + usually an extension
  - customers_info.txt (Name: customers_info and Extension: .txt)
  - data.tar.gz   (extension .tar.gz)
  - data (no extension)
- Directory (Folder): where the files exist
- Full Path: The whole path to access a file on the OS
  - /home/moustafa/workspaces/myfile.txt   [unix]
  - C:\\moustafa\\workspaces\\myfile.txt   [windows]
- Relative path
  - Relative to a current directory. Let's say you are at /home/moustafa
    - Then on relative path is **workspaces**/myfile.txt   [or ./**workspaces**/myfile.txt]
  - Our code refers typically to relative, never to full path. [your code will be moved]

. and ..

- . refers to the current directory
- .. refers to the parent director
- Let's say we have the following path
- code/cpp_skills/z/../../cpp_skills/z/y/mydir/x
  - .. means move up level so let's simulate
  - code/cpp_skills/z/../../cpp_skills/z/y/mydir   [cancel z/..]
  - code/cpp_skills/../cpp_skills/z/y/mydir   [cancel cpp_skills/..]
  - code/cpp_skills/z/y/mydir
- You should always be careful about path with spaces
  - In commands, in project workspaces, etc
  - Use "" for CL commands

Line Endings

- Not straightforward due to historical reasons
- Reading and writing may be different from an Operating system to another
  - Careful for moving files between windows and unix
  - Don't be shocked if moved files and opened them to find them displayed differently!
- Info
  - Carriage Return (CR or \r)
  - Line Feed (LF or \n)
- In windows lines end uses **\r\n**
- In Unix and the newer Mac use **\n**

Character Encodings

- It is used to represent a repertoire of characters by some kind of encoding system that assigns a **number to each character** for digital representation
- Common encodings:
  - ASCII: only 128 characters
  - UNICODE: designed to **cover all** the world's major living **languages**
    - UTF-8: Very popular
    - UTF-16: Much bigger, but less used in practice, so far
- Opening a file with a lower encoding capabilities, you will face problem
  - E.g. opening UTF-8 file using ASCII
- In python: The default encoding is **platform dependent**
  - So be careful. If you should use UTF-8, then pass it as parameter

UTF-8 File content

Dürst. ˈmaˈkʊs

Σ f(i) = Π g(i), $\forall$x$\in$$\mathbb{R}$: ⌈x⌉ = −⌊−x⌋, α ∧ ¬β = ¬(¬α ∨ β),

2H₂ + O₂ ⇌ 2H₂O, R = 4.7 kΩ, ø 200 mm

ði ɪntəˈnæʃənəl fəˈnɛtɪk əsoʊsiˈeɪʃn

Σὲ γνωρίζω ἀπὸ τὴν κόψη

ὄταν τ΄ εἰς τὰ πράγματα ἀποβλέψω καὶ ὄταν πρὸς τοὺς

გთხოვთ ახლავე გაიაროთ რეგისტრაცია Unicode-ის მეათე საერთაშორისო

Зарегистрируйтесь сейчас на Десятую Международную Конференцию по

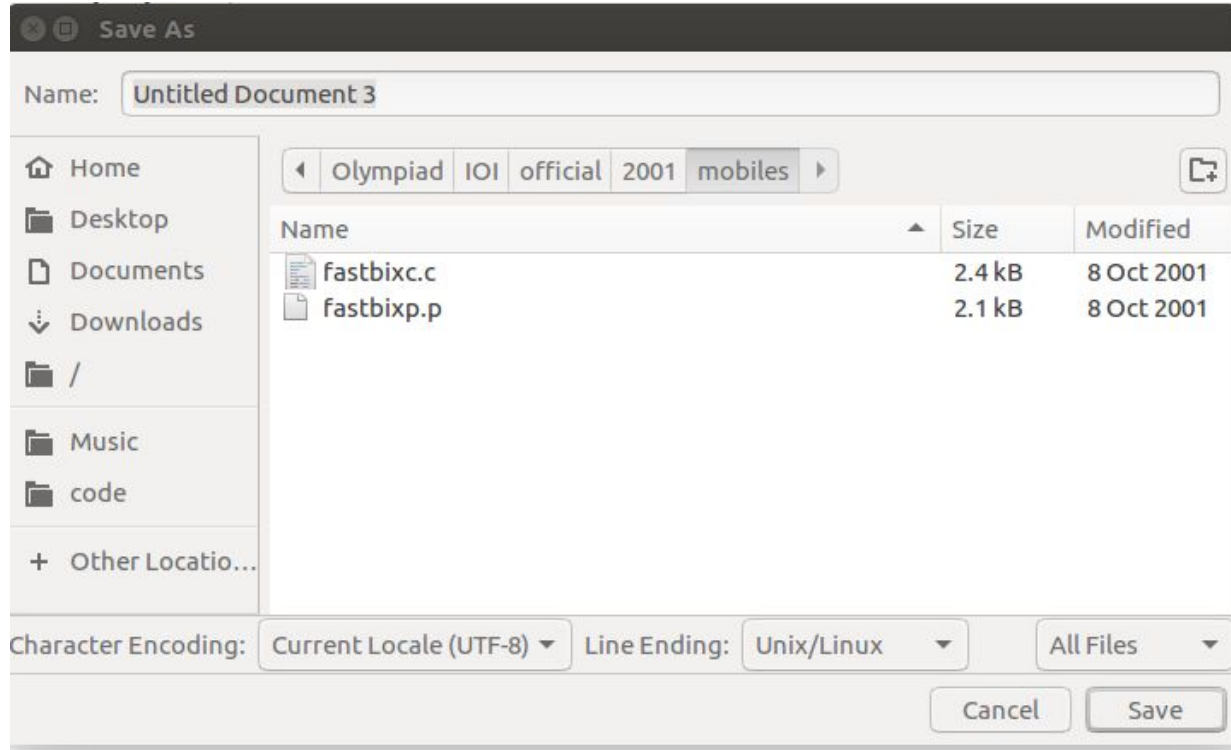◉ แผ่นดินฮั่นเสื่อมโทรมแสนสังเวช พระปกเกศกองบู๊กู้ขึ้นใหม่

ሰማይ አይታረስ ንጉሥ አይከሰስ።

ᚾᛗ ᛚᛈᚱᛈ ᚦᛈᛏ ᚾᛗ ᛒᚾᛊᛗ ᚠᛏ ᚦᛈᛗ ᛚᚠᛏᛊᛗ ᛏᚠᚱᚦᛈᛗᚠᚱᛊᛗ ᛈᛁᚦ ᚦᚠ ᛈᛗᛒᛈ

⠠⠆⠊⠲⠂⠄⠂⠶⠀⠠⠂⠆⠆⠂⠄⠒⠦⠀⠠⠂⠠⠂⠄⠂

Saving file in Ubuntu

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."