Python 101

Working with Data Formats (Texts and Multimedia) and Pickling

Contents

1. Base64
2. JSON
3. CSV
4. XML
5. XLSX
6. YAML
7. Pickle
8. A/V/I (PyAudio, Scikit-Video, Pillow)

Base64

Base64 is a way of representing binary data. It is a way of binary to text encoding.

It consists of 0-63 values each represented with A-Z,a-z,0-9, + , /

It is often used to transfer binary data over media where only textual data is meant to be transferred eg. Transferring Images and other media over HTTP.

In Python3 we use **base64** module for encoding and decoding base64.

The module also provides b16, b32, and b85 encodings but b64 is the norm for transferring binary over textual streams.

JSON

JSON Stands for Javascript Object Notation. It is literally the same thing. Its the way an object can be denoted in Javascript.

JSON can be created with the following notations-

{ key: value } = a comma seperated key value pair

[ ] = a comma separated values only

The keys can only be String and values can be of type String, int, float, or boolean.

It is the most preferable way of transfering data with description.

CSV

Comma Separated Values or CSV are a set of a delimeter separated values and it is most commonly used for import and export of dbs and spreadsheets.

Mind that delimeter mostly is comma but can be any other character. Many implementations prefer using semi-colon instead of comma as it doesn’t collide with number , address representations etc.

In Python we can read and write csv files using **csv** module.

XML

e**X**tensible Markup Language is one of the most popular (after JSON) format to share data.

XML is said to be self descriptive.

SOAP (Simple Object Access Protocol) uses XML to communicate with other instances of OSes through HTTP.

XML is quite similar to that of HTML as both are subsets of SGML (Standard Generalized Markup Language). It is implemented using **lxml** in python.

https://lxml.de/tutorial.html#the-element-class

XLSX

XLSX files, More formally known as Excel files are a very common way of dealing with the spreadsheets data. Making any backend related to ERP might require parsing Excel files.

In Python **xlwt** and **xlrd** (external) are used for this purpose.

https://github.com/python-excel

YAML

YAML stands for YAML ain't markup language its a recursive abbreviation just like GNU. It is a data serialization language focused on being readable.It is not used for data exchange unlike XLSX, XML, CSV, or JSON.It is best suited for creating configuration files. It's main focus is at being human readable.

JSON lacks a big thing in terms of human readability and that is comments which YAML supports. Apart from this YAML also supports self referencing and support for complex data structures. It is used with **yaml (external**)module of python.

https://learnxinyminutes.com/docs/yaml/

Pickle

Pickling is the serialization and Deserialization of Python objects into bytes to provide them persistence in Secondary storage.

There are two methods namely **load** and **dump** under **pickle module** which work identically to json’s load and dump.

Audio Video and Image R/W in Python

Working with Multimedia is extremely easy in Python but knowing the right tools is the most important part here. Each tool is worth of a whole courseware in itself but a general i/o is essential to know.

Audio - wave

Images - PIL (Python Imaging Library)

Video - Scikit-Video or OpenCV (Scikit Video is enough if the purpose is small)

We should avoid using OpenCV unless the task involves sophisticated algorithm usage as its a large library.

Audio Video and Image R/W in Python contd..

The most appropriate to process types in AVI are-

Audio - wav (waveform audio file) - **ffmpeg** can be used to convert other formats to wav

Images - png / jpeg

Videos - mp4