

HELLO EVERYONE
This is Shuno.

And today,,, I decided to download the Entire “The Office” show, all seasons, all episodes, on my laptop.

To enjoy free buffer-free smooth streaming even when I exhaust data limit or even when web server streaming is very slow.

So,,, I searched up the internet and I found this nice sweet little directory which contains all of “THE OFFICE” episodes in an organized manner :D

So this, as you can see, is the link:
<http://51.158.153.210/the-office/>

and well, I started downloading the episodes one by one starting from S01 which, understandably is a VERY SLOW AND REPETITIVE process :(

The surprising fact was that each episode is hardly 60 – 80 MB (which is pretty small considering webseries episodes these days which are easily 200+ MBs.

SO!!! After manually downloading full S1 and then about 10 episodes of S2 one by one (I cant’s even download 2 at once since in doing so, speed goes down terribly!!!!) I finally decided!

FUCK THIS SHIT

It’s easier to write a code to automate this download one-by-one task >:(
I feel like a dumb robot wasting 2 hours already and there’s 7 whole seasons more left to download :(

SOOOOOO I STARTED CODING ;D
heheheheh

the first google search I did was this:
‘python code to download videos one by one from online directory’

and it took me to the first site:
<https://www.geeksforgeeks.org/downloading-files-web-using-python/>

so, I scrolled down, and there it was! A code to BATCH DOWNLOAD videos from an online directory :DDD

however, just to make sure this code would be of my use (since, yk, I dont know how online directories and websites and html really work,) I decided to open this url “directory” on which this code was working on...

<http://www-personal.umich.edu/~csev/books/py4inf/media/>

however, to my disappointment,,,,,
THIS SITE DID NOT EXIST :(((
(big sad)

anyways,,, my next step:
I went on to understand the code and how it worked. :D

so I copy pasted this skeleton code, onto my VS Code workspace and started tinkering :D

so, as you can see (you actually cannot, sorry haha xp) I actually quite understood the code :O!

This part was a function for extracting all the hyperlinks with a .mp4 extension from the given input html. This function returned a list of video .mp4 hyperlinks which we want to download.

Now, this function is for opening these links one by one and downloading the video, chunk by chunk (download IS basically reading chunks of bytes from the server and writing it into your own PC, sounds kinda WOAHH-type if you think about it deeply eh?)

ok so, yeah, this function downloads the videos from the list one by one :D

hmmmmmmmmmm

so, I decided! Hey, lemme change a few things to download the entire office from S3 to S9 (cuz I was already in the midst of downloading S2 manually, so i'd finish it manually and leave the rest to auto)

So I made (actually, MODIFIED, cuz that's what coders do B)) the code to do the following:

1. iterate a variable x from 3 to 9 (S3 to S9)
2. create a folder named Sx (if it doesn't exist, using os.mkdir)
3. now, for each season (x), run the above download code to obtain the videos list present in season x of the main directory
4. for each video of season x from obtained list, save (download) the video into the folder Sx

so, yeah, this code would help ME auto download S3 to S9 all videos neatly into my computer, in individual folders names S3, S4, S5, ...

BUTTTTT then I thought, hey! Why not make a code which anyone can use???

so I made further modifications:

5. made two variables start_season and end_season which the user can adjust IN the code itself (I have not done error handling yet, so hopefully the user is intelligent enough to keep it from 1 to 9 haha)

6. now, since x iterates from 1 to 9, I have made an if condition,,, that if the video already exists in our Sx directory, then it will skip this one and continue downloading the other videos in that season....

SOOOO NOW THE FINAL MOMENT
enough explanation!

LET US PRESS THE BUTTON AND RUN THE CODE!!! >:D
(I shall upload this code on github if it works, ;') it's amazing how the need to download The Office has made me learn a bit more about coding ah :~D just like how that 4chan user solved the superpermutation problem just to watch an anime xD, ah, the need, the need! Necessity is the mother of inventions, true, true,,

NOW ENOUGH TALK.
Time to Execute Order 66 >:)

I sure hope my laptop doesnt crash :~(

oh god the code is running!!!!
lemme check the folder :O

ooh! It did download :O!!!!
OMG
AND IT LEGIT RUNS :~DDD!!!!

YASSS CODE IS RUNNING LETS GOOOOOOOOO

anyways, on a side note,,, I had noticed this while modifying this code and that was,,, originally, in the HTML sourcecode of these video files,,, the urls appeared like this:

<http://51.158.153.210/the-office/season-2/The%20Office%20S02E04%20The%20Fire.mp4>

so, as you can see,,, what is supposed to be a space bar (' ') is replaced by a '%20' which is quite understandable since, in terms of real coding, space bars are LETHAL. Never use space in naming any main directory. Always use underscores instead. So, 'The_Office' instead of 'The Office'. (but please excuse my this time since I have indeed used the latter this time haha ^^;) BUT COMING BACK TO THE POINT

<http://51.158.153.210/the-office/season-2/The Office S02E04 The Fire.mp4>

replacing the '%20' with spaces (' ') was what the real name of the files were. So, in the code, I had done exactly that. I had replaced them implicitly in the code, while saving the filename. As you can see here:

```
file_name = file_name.replace('%20', ' ')
see?
```

But well, it seems I might have failed to consider other punctuations than spaces(%20)....

As in S02E16....

The Office S02E16 Valentine's Day.mp4 (this is the original file name on downloading manually)

The Office S02E16 Valentine%27s Day.mp4 (and this is the filename that was downloaded by the code...)

so you see, there's also an apostrophe(''), which is written as '%27' and needs to be replaces in the code...

SO as I have said, the code needs a bit more tweaking.... (another example is just the next episode:

The Office S02E17 Dwight%27s Speech.mp4

so yeah. I shall fix these....

BUTTTTT

THE CODE IS WORKING :”DDDDDDDD!!!!!!

OH LOOK HOW THE COMPUTER IS AUTO DOWNLOADING THE VIDEOS :”DDDDDD

and mostly, it works :”DDDDDDDD

heh, so yes.

In conclusion, Shuno is GENIUS >:”3

(I shall now end this video, thank you for watching. I hope I can add some cool songs or stuff into it to make it less boring. And also, I shall speed up this video in case it is too long which it is.,,,)

I shall now let the code run and let it download as much as it can, after which, in case the data gets exhausted, download will not proceed further and i'll have to ctrl+C from the terminal to stop the program. And well, rerun it.

So as I said, I shall make modifications to also allow for code to automatically stop once data transfer rate drops to very low for some while...

If you have any more ideas, feel free to comment :D

Like, Subscribe and Share! :D

xD

LMAO just kidding who'll watch this anyways?

Anyways, github link in description :)

Thank you fellow Office Fans,
Kudos to you,

and special thanks to my beloved friend,
Sis :)

(she's the one who got me into The Office and awwhh,, I want what Jim and Pam have :))

GOODBYE!!!!!!

endnotes:

yes, my data got exhausted, and code got stuck at S03E13... I stopped the code, using the classic old ctrl+C on terminal,,, and on opening the half downloaded video, it did open, and it ran pretty well, upto 11:50 after which it stopped abruptly.

In short, the code works and I am happy :))

it is Sunday, Dec 11, 01:45 AM in the night and I am a happily screeching Office fan amazed at my brilliant sense of mine and coding skills and most of all, thankful for the internet, because 90% of the code's praise (and in fact the code itself) goes to online, the internet, and GFG :)"D so yes. Thank you!