

Web scraping youtube songs from insomnia.gr & selenium bot that adds songs to youtube playlist

Εισαγωγή

Στο insomnia.gr υπάρχει ένα forum που ανεβάζει κόσμος το ποιο τραγούδι ακούει τώρα. Πάνω από 400 σελίδες με καμιά ντουζίνα τραγούδια περίπου η κάθε μία. Ένας φίλος μου ήθελε να φτιάξει μια playlist με τα τραγούδια από εκεί, οπότε στην αρχή έφτιαξα ένα script που scrapάρει όλες τις σελίδες.

Παρατήρησα ότι τα link για την κάθε σελίδα ήταν παρομοιότυπα:

<https://www.insomnia.gr/forums/topic/695707-%CF%80%CE%BF%CE%B9o-%CF%84%CF%81%CE%B1%CE%B3%CE%BF%CF%8D%CE%B4%CE%B9-%CE%B1%CE%BA%CE%BF%CF%8D%CF%84%CE%B5-%CF%84%CF%8E%CF%81%CE%B1/page/ + Αριθμός Σελίδας.>

Οπότε αρχικά έτρεξα μια λούπα για την κάθε μία σελίδα όπου ψάχνει να βρει όλα τα youtube links.

Σημείωση 1:

Κοιτάω μέσα μόνο για συγκεκριμένο τρόπο που μπορεί να εμφανιστεί το link

```
new_songs=set(re.findall(r'https://www.youtube.com/embed/([\w\-\_]*)',response.text))
```

δηλαδή ο κώδικας, συγκεκριμένα το Regular expression που χρησιμοποιώ δεν βρίσκει όλα τα format ενός youtube link (πχ. <http://youtu.be/-wtIMTCHWuI>). Τα links που βρήκα τα πρόσθεσα σε μια “playlist” και μετά τα πρόσθεσα στην “final_list ένα ένα στην περίπτωση που δεν υπήρχαν ήδη. Στην συνέχεια τα αποθήκευσα σε ένα xlsx αρχείο.

Σημείωση 2:

Για playlist λίγων τραγουδιών δουλεύει ο κώδικας που είναι σχόλια. Για λίστα με εκατοντάδες τραγούδια προφανώς και όχι. Επίσης η playlist που δημιουργείται είναι unlisted.

Οπότε πήγαμε στην δεύτερη φάση του project να φτιάξω ένα botaki που θα τα προσθέτει ένα ένα όλα τα κομμάτια σε μια playlist.

Το αρχικό πρόβλημα που αντιμετώπισα ήταν ότι έπρεπε να μην ανοίγει νέα καρτέλα του chrome αλλά να δουλεύει σε ήδη υπάρχουσα όπου υπάρχει συνδεδεμένος λογαριασμός στο youtube για να μπορεί να δημιουργηθεί η Playlist και να προσθέσει τα τραγούδια.

Αυτό επιτεύχθηκε με το να τρέχω στο φάκελο όπου βρίσκεται το chrome εντολή `chrome.exe –debugging-port = PORT_NUMBER` και να πειράζω τα options του webdriver.Chrome.

Στην συνέχεια τρέχω το script που κοιτάζει για κάθε ένα link στο αρχείο που αποθηκεύσαμε πιο πριν αν είναι available και αν είανι να πατάει το κουμπί save και μετά να το κάνει προσθήκη στην λίστα insomnia1.

\m/

Tsaparov John