

2020-2021

Ατομική έκθεση Python

Δημοσθένης Καραμπάρπας

Η python αν και είχα επαφή για πρώτη φορά με αυτήν σε αυτό το μάθημα μου κίνησε το ενδιαφέρον και για αυτό δέχτηκα ένα αξιοπρεπές κομμάτι της εργασίας να είναι προσωπική μου ευθύνη. Πιο συγκεκριμένα στράφηκα προς το κομμάτι του διαδικτύου γράφοντας πρακτικά το κομμάτι του internet_part. Ο σκοπός αυτού του κομματιού είναι να παίρνει τα στοιχεία από την ιστοσελίδα nemertes και να οργανώνει τα στοιχεία ανά μαθητή.

Αρχικά, το όραμα για το πως λειτουργεί το πρόγραμμα ήταν να μετακινείται στις σελίδες, παίρνοντας τα hyperlink που αντιστοιχούν σε κάθε μαθητή, να τα ανοίγει και από εκείνη την σελίδα να αντλεί τις πληροφορίες του κάθε μαθητή και το link του pdf. Αυτό αρχικά θα γινόταν με τη χρήση των modules request και bs4 αλλά λόγω προβλημάτων με τα certification της ιστοσελίδας και ενημέρωση από άλλη ομάδα με παρόμοιο θέμα, η request αντικαταστάθηκε από την urllib.request.

Το πρόγραμμα στη συνέχεια γρήγορα αναπτύχθηκε χωρίς ιδιαίτερες δυσκολίες δημιουργώντας μία συνάρτηση για τις κεντρικές σελίδες που έπαιρνε το hyperlink ανά μαθητή, μία συνάρτηση για τις δευτερεύουσες που αντλούσε τις πληροφορίες από τον κώδικα html των hyperlinks και μια κλάση στη οποία αποθηκεύονταν οι πληροφορίες κάθε μαθητή. Αυτές οι κλάσεις έπειτα έμπαιναν σε μία list.

Έπειτα υπήρχαν δύο bugs τα οποία ήταν ότι τα keywords της ιστοσελίδας ήταν χωρίς κενά πιο διορθώθηκε εύκολα μετά από την καλύτερη κατανόηση της get.text() και το ότι μερικές διπλωματικές δεν είχαν όλες τις πληροφορίες που ήθελα, κάτι που δημιουργούσε λάθος στο dictionary που διορθώθηκε με την χρήση των στοιχείων του site για keys αντί για απλή αρίθμηση.

Τέλος, πραγματοποιήθηκαν μικρές αλλαγές για την καλύτερη σύνδεση με άλλα κομμάτια του κώδικα.

Όρες εργασίας:

Όλη η διαδικασία πήρε περίπου 15 ώρες δουλειάς στην οποία περίπου οι 10-11 ήταν η δημιουργία του κυρίως προγράμματος και οι υπόλοιπες τα bugs. Οι τελικές μικρές αλλαγές ήταν χρονικά ασήμαντες.

Πηγές:

Νικόλαος Αβούρης, Μιχαήλ Κουκιάς, Βασίλειος Παλιουράς, Κυριάκος Σγάρμπας,

“Python Εισαγωγή στους υπολογιστές” , Πανεπιστημιακές Εκδόσεις Κρήτης,

ISBN 978-960-524-529-0

<https://www.datacamp.com/community/tutorials/pip-python-package-manager>

<https://stackoverflow.com/questions/7243750/download-file-from-web-in-python-3>

https://brightdata.com/blog/how-tos/how-to-use-beautiful-soup-for-web-scraping-with-python?kw=&cpn=14745430544&cam=aw_blog_dynamic_547760284460&utm_term=&utm_campaign=blog&utm_source=adwords&utm_medium=ppc&utm_content=dynamic&hsa_acc=1393175403&hsa_cam=14745430544&hsa_grp=131242020607&hsa_ad=547760284460&hsa_src=g&hsa_tgt=dsa-39587879683&hsa_kw=&hsa_mt=&hsa_net=adwords&hsa_ver=3&gclid=CjwKCAiAtouOBhA6EiwA2nLKH_AaZ2RlupL0dpZhlMV-Gf-ZF1QzlcaW8g-iUKbTTFSWlbPVbvWGBBoCuh0QAvD_BwE

<https://stackoverflow.com/questions/22726860/beautifulsoup-webscraping-find-all-finding-exact-match>

<https://www.guru99.com/accessing-internet-data-with-python.html>

<https://stackoverflow.com/questions/6287529/how-to-find-children-of-nodes-using-beautifulsoup>

<https://stackoverflow.com/questions/1080411/retrieve-links-from-web-page-using-python-and-beautifulsoup>

<https://www.geeksforgeeks.org/downloading-pdfs-with-python-using-requests-and-beautifulsoup/>

<https://www.codegrepper.com/code-examples/python/download+pdf+from+link+using+python>

<https://stackoverflow.com/questions/8933863/how-to-find-tags-with-only-certain-attributes-beautifulsoup>