

Μεταγλωττιστές 2020

Προγραμματιστική Εργασία #2

Ονοματεπώνυμο: ΦΟΥΡΝΑΡΗΣ ΓΙΩΡΓΟΣ

A.M.: Π2017158

1. ΣΥΝΟΠΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΒΗΜΑΤΩΝ

Αρχικά έκανα import την βιβλιοθήκη re στο πρόγραμμά μου και έπειτα ξεκίνησα να λύνω ένα ένα τα ερωτήματα που μου είχαν ανατεθεί, αφού πρώτα είχα δημιουργήσει μια κανονική έκφραση και είχα ανοίξει το αρχείο μου.

2. ΠΕΡΙΓΡΑΦΗ ΚΑΝΟΝΙΚΗΣ ΕΚΦΡΑΣΗΣ ΣΕ ΚΑΘΕ ΒΗΜΑ

1. `<title>(.*?)</title>`

Επιλογή των χαρακτήρων που βρίσκονται μεταξύ τους με την χρήση των (.) και (+).

2. `<!--.*?-->,re.DOTALL`

Επιλογή των σχολίων και απαλοιφή τους.Χρήση * εφόσον δεν χρειάζεται να υπάρχει τουλάχιστον ένας χαρακτήρας μεταξύ τους.Χρήση re.DOTALL για σχόλια με παραπάνω από μια γραμμές.

3. `<(script|style).*?*</(script|style)>,re.DOTALL`

Απαλοιφή όλων των χαρακτήρων μεταξύ script και style.Χρήση | για να συμπεριλαμβάνονται και τα δύο.

4. `<a.*?href="(.*?)">(.*?),re.DOTALL`

Εξαγωγή περιεχομένων του href και του a.

5. `<[>]+>,re.DOTALL`

Απαλοιφή όλων των tags.

6. `&(amp|gt|lt|nbsp);`

Μετατροπή των HTML entities.

7. `\s+`

Απαλοιφή των whitespaces μια ή περισσότερες φορές.

3. ΠΗΓΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΣΑ

Χρησιμοποίησα το υλικό που υπάρχει στο site του μαθήματος.