**Μεταγλωτιστές 2020**
**Προγραμματιστική Εργασία #2**

**Ονοματεπώνυμο :** Τιμόθεος Αυγερινός
**ΑΜ :**Π2015009

## 1. Συνοπτική περιγραφή της σειράς βημάτων επεξεργασίας στον κώδικά σας

Έγιναν με την σειρά που ζητήθηκαν όλα τα ερωτήματα χρησιμοποιώντας οχτώ(8) κανονικές εκφράσεις . Έπειτα με την **with** άνοιξα το αρχείο **testpage.txt** και έλεγξα το output αρχείο που παράχθηκε από το πρόγραμμα.

## 2. Περιγραφή της κανονικής έκφρασης που χρησιμοποιήσατε σε κάθε βημα

**Ερώτημα #1**

**(r'<title>(.+?)</title>')** : Επιλογή ενός ή παραπάνω χαρακτήρα που βρίσκετε μέσα στο title tag. Χρησιμοποιήθηκαν οι τελεστές **.** (οποιοσδήποτε χαρακτήρας) και **+** (μια ή περισσότερες φόρες ο χαρακτήρας που προηγείται) και συγκρατούνται με τις παρενθέσεις.

**Ερώτημα #2**

**(r'<!--.*?-->',re.DOTALL)** : Για την απαλοιφή των σχολίων χρησιμοποιήθηκε ο τελεστής * (0 ή περισσότερες φορές ) διότι μπορεί να υπάρχει κενό σχόλιο και επίσης χρησιμοποιήθηκε και το re.DOTALL επειδή μπορεί να χουμε σχόλια πολλαπλών γραμμών .

**Ερώτημα #3**

**(r'<(script|style).*?>.*?</(script|style)>',re.DOTALL) :** Για την επιλογή όλων των Script και style tags χρησιμοποιήθηκε ο τελεστής **|** για να επιλέγονται και τα δυο καθώς και οι τελεστές **.** Και ***** .

**Ερώτημα #4**

**(r'<a.+?href="(.*?)".*?>(.*?)</a>',re.DOTALL) :** Για την εξαγωγή όλων των περιεχομένων του **href** και του περιεχομένου του **a** χρησιμοποιήθηκε η παραπάνω έκφραση.

**Ερώτημα #5**

**(r'<.+?>|</.+?>',re.DOTALL) , (r'<.+?/>',re.DOTALL) :** Χρησιμοποιούνται δύο κανονικές εκφράσεις γιατί ένα tag μπορεί να είναι **self-closing** . Με αυτούς τους 2 τρόπους γίνετε ταίριασμα των **<></>**.

**Ερώτημα #6**

**(r'&(amp|gt|lt|nbsp);') :** Εξαγωγή των html entities .

**Ερώτημα #6**

**(r'\s+') : Εξαγωγή** των **whitespaces**(\s) με τον τελεστή **+** (μία ή περισσότερες φορές).

## 3. Αναφορά σε πήγες που πιθανόν χρησιμοποιήσατε

Οι πήγες ήταν οι σημειώσεις του μαθήματος που βρίσκονται εδώ :http://mixstef.github.io/courses/compilers/lecturedoc/unit2/module1.html#sub

Και στις εργαστηριακές ασκήσεις του μαθήματος.