Μεταγλωτιστές 2020
Προγραμματιστική Εργασία #2

Οναματεπώνυμο : Αφεντούλης Κωνσταντίνος
ΑΜ : Π2015021

1. Όπως αναγράφεται και στα σχόλια του κώδικα στο αρχείο **html-processor.py,** χρησιμοποιήσαμε κανονικές εκφράσεις για την υλοποίηση των ερωτημάτων (κάθε βήμα αντιστοιχεί σε ένα από τα βήματα της εκφώνησης). Η εισαγωγή του αρχείου **testpage.txt** έγινε με την εντολή **with** και με το τρέξιμο του προγράμματος πήραμε το **output**.

2.

ΒΗΜΑ 1

(r'<title>(.+?)</title>') : Με την χρήση του τελεστή **.** επιλέγεται οποιοσδήποτε χαρακτήρας, ενώ με το + επιλέγεται μια ή περισσότερες φορές ο χαρακτήρας που προηγήθηκε. Με την χρήση των **( )** κρατάμε το περιεχόμενο τους το οποίο και τυπώνεται στην συνέχεια του κώδικα.

ΒΗΜΑ 2

(r'<!.*?>',re.DOTALL) : Με την χρήση του τελεστή ***** στην απαλοιφή των σχολίων σιγουρευόμαστε οτι συμπεριλαμβάνονται και τα “κενά” σχόλια (0 ή περισσότερες φορές). Χρήση του **re.DOTALL** για ύπαρξη σχολίων πολλαπλών γραμμών.

ΒΗΜΑ 3

(r'(<script.+?</script>)|(<style.+?</style>)') : Επιλέγει ό,τι βρίσκεται μέσα στα

ΒΗΜΑ 4

('<a.*?href="(.+?)".*?>(.*?)',re.DOTALL) : Εξαγωγή περιεχομένων **href** και **a.**

ΒΗΜΑ 5

(r'<.+?>|</.+?>',re.DOTALL), (r'<.+?/>',re.DOTALL) : Δύο κανονικές εκφράσεις για self-closing tags.

ΒΗΜΑ 6

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

ΒΗΜΑ 7

(r'\s+') : Χρήση +(μια ή περισσοτερες φορές) για εξαγωγή **whitespaces(\s)**

3. Χρησιμοποιήθηκε υλικό του μαθήματος από:
1. https://gist.github.com/mixstef/39d5257c7498dceac1aa6428e33f2003#file-s050-sub-callback-py
2. https://gist.github.com/mixstef/39d5257c7498dceac1aa6428e33f2003#file-s010-hint-keep-only-words-py
3. http://mixstef.github.io/courses/compilers/lecturedoc/appendix-python/module1.html#id5
4. http://mixstef.githourub.io/cses/compilers/lecturedoc/unit2/module1.html#id8
5. http://mixstef.github.io/courses/compilers/lecturedoc/unit2/module1.html#sub