

Μεταγλωτιστές 2020

Προγραμματιστική Εργασία 2

Ονοματεπώνυμο: Αντρέας Παππούτας
ΑΜ: Π2017148

Αρχικά σε αυτό το πρόγραμμα έκανα import το re της python. Ακολούθως δημιουργώ τις μηχανές ταιριάσματος και τις κανονικές εκφράσεις. Επιπλέον έφτιαξα μια συνάρτηση για τη αντικατάσταση & > < & nbsp. Τέλος διαβάζουμε το testpage.txt, κάνουμε χρήση των εκφράσεων και της συνάρτησης και τα εκτυπώνουμε.

Ερώτημα 1:

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

Σε αυτό το ερώτημα μέσω της . μαζί με + γίνεται ταίριασμα των χαρακτήρων που είναι ανάμεσα στο <title></title> και με τις παρένθεσης η συγκράτησης τους.

Ερώτημα 2:

`('<!--.*?->',re.DOTALL)`

Η χρήση της πιο πάνω έκφρασης είναι η απαλοιφή των σχολίων. Για να γίνει απαλοιφή ολοκλήρων των σχολίων χρησιμοποίησα group(0).

Ερώτημα 3:

`(r'<(s(?:cript|tyle)).*?>.*?</1>',re.DOTALL)`

Χρησιμοποιώντας . μαζί με * ταιριάζω ότι έχει μέσα στο script και style. Επιπλέον κάνω χρήση του /1 για ότι έχει στο group(1).

Ερώτημα 4:

`(r'<a.+?href="(.*?)".*?>(.*?)',re.DOTALL)`

Η πιο πάνω έκφραση ταιριάζει το περιεχόμενο του href και του a.

Ερώτημα 5:

`(r'<.+?>|</.+?>',re.DOTALL)`

`(r'<.+?>',re.DOTALL)`

Με τις 2 εκφράσεις πιο πάνω γίνεται το ταίριασμα των <></>.

Ερώτημα 6:

`(r'&(amp|gt|lt|nbsp);')`

Αυτή η έκφραση έγινε για τα & > < & nbsp.

Ερώτημα 7:

`(r'\s+')`

Μέσω αυτής της έκφρασης γίνεται ταίριασμα των whitespace.