

# Μεταγλωτιστές 2020

## Προγραμματιστική Εργασία #2

Ονοματεπώνυμο : Σιδεράς Γεώργιος

AM: Π2017164

Αρχικά κάνουμε import την βιβλιοθήκη re που είναι υπεύθυνη για τις regular expressions , έπειτα φτιάχνω μια συνάρτηση που θα χρησιμοποιηθεί στην συνέχεια. Δημιουργούμε τις διάφορες εκφράσεις και αναλυτικότερα :

- titlos) κάτι που να βρίσκεται μεταξύ <title> και οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές ακολουθούμενο απο </title>
- scholia) κάτι που να βρίσκεται μεταξύ <!-- και οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές ακολουθούμενο απο →
- other\_tags) κάτι που να βρίσκεται μεταξύ <script> ή του <style> και οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές ακολουθούμενο απο είτε </script> είτε </styles>
- href) κάτι που να βρίσκεται μεταξύ <a και οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές ακολουθούμενο απο href="" μέσα στα αυτάκια οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές , μετά τα αυτάκια πάλι οποιοσδήποτε χαρακτήρες πολλές φορές ενα σύμβολο > πάλι οποιοσδήποτε χαρακτήρας και τέλος τους χαρακτήρες </a>
- tags) κάτι που να βρίσκεται μεταξύ < και οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές ακολουθούμενο απο >
- cntitics) κάτι που να βρίσκεται μεταξύ & και οποιονδήποτε χαρακτήρα που μπορεί να επαναληφθεί πολλές φορές ακολουθούμενο απο τον χαρακτήρα ;
- many\_spaces) κάτι που να έχει παραπάνω απο μια φορά το κενό χαρακτήρα.

Ο λόγος που χρησιμοποιήσαμε την συνάρτηση re.DOTALL μερικές φορές είναι οτι το ταίριασμα με την τελεία (.) δεν διασχίζει πολλαπλές γραμμές και με την βοήθεια αυτής της συνάρτησης το κάνει.

Στην συνέχεια ανοίγουμε ένα αρχείο εισόδου ονόματι "testpage.txt" με δικαιώματα read only και του δίνουμε όνομα fr το οποίο στην επόμενη σειρά το διαβάζουμε όλο και το γράφουμε σε μια μεταβλητή text. Έπειτα ανοίγουμε ένα άλλο αρχείο με όνομα output.txt με δικαιώματα write ώστε να γίνει το αρχείο εξόδου μας.

Στα επόμενα βήματα απλά γράφουμε στην μεταβλητή `text` όλες τις αλλαγές απο τα αρχικό κείμενο με βάση των κανονικών εκφράσεων και στο τέλος εκτυπώνουμε το περιεχόμενο της στο αρχείο εξόδου. Επίσης χρησιμοποιούμε την μέθοδο `sub` που αντικαθιστά οτι του δώσαμε ως κανονική έκφραση με ένα κενό χαρακτήρα , και στην περίπτωση της έκφρασης `cntitits` με την συνάρτηση `subl` όπου ανάλογα αν ανάμεσα απο το "&" και το ";" βρίσκεται το "amp" ή το "gt" ή το "lt" ή το "nbsp" με το ανάλογα σύμβολα "&" ">" "<" "(κενός χαρακτήρας)".

Οι πηγές που χρησιμοποίησα για να κάνω την άσκηση ήταν οι online διαφάνειες του site :  
"<http://mixstef.github.io/courses/compilers/lecturedoc/unit2/module1.html>".