

Μεταγλωττιστές 2020

Προγραμματιστική Εργασία 2

Ονοματεπώνυμο: Σπύρος Μπαλωμένος

ΑΜ: Π2017032

Περιγραφή κανονικών εκφράσεων

- i. Για την εκτύπωση του τίτλου χρησιμοποιείται η κανονική έκφραση `<title>(.*?)</title>` όπου αναγνωρίζει οτιδήποτε βρίσκεται μεταξύ των tags και με την εκτύπωση του group(1) εμφανίζεται ο τίτλος.
- ii. Για την απαλοιφή των σχολίων χρησιμοποιείται η κανονική έκφραση `<!--.*?-->` όπου μέσω της sub διαγράφει τα σχολια της HTML.
- iii. Η κανονική έκφραση `<script>(.*?)</script>|style=(.*?)</style>` θα διαγράψει τα scripts και styles (δεν υπήρχαν στην συγκεκριμένη περίπτωση).
- iv. Η κανονική έκφραση `'<a(.*?)'` θα αναγνωρίσει όλο το περιεχόμενο μεταξύ των a Tags .
- v. Η επόμενη κανονική έκφραση `'<.*?>(.*?)</.*?>` εντοπίζει όλα τα είδη των tags και τα διαγράφει με την sub.
- vi. Η χρήση της έκφρασης `'(&|>|<|);'` εντοπίζει τα entities στο περιεχόμενο και τα αντικαθιστά αναλόγως.
- vii. Τέλος η κανονική έκφραση `\s+` αναγνωρίζονται συνεχόμενα spaces καθώς και tabs και απαλείφονται μέσω της sub.