

## Περιγραφή Σειράς Βημάτων Επεξεργασίας του Κώδικα

Πρώτα ανοίγω το αρχείο και διαβάζω τον html κωδικα και τον βάζω σε δυο μεταβλητες text και newText. Στο test δοκιμαστικά θα πηγαίνω κάθε φορά και θα εκτελώ μια re και θα αποθηκευω το αποτέλεσμα του σε ένα αρχείο. Στην συνέχεια, θα εκτελω το re στο newText. Το αποτελεσμα του newText θα αποθηκευεται στο newText και ετσι θα δημιουργήσω το αποτελεσμα που έχει περάσει από όλες τις κανονικές εκφράσεις. Τελος, αποθηκεύω το newText στο αρχείο output.txt οπως ζητηθηκε. Τα αρχεία που γίνονται μια μια οι κανονικες εκφράσεις έχουν συμπεριληφθεί εντος του φακέλου 'myouts' καθώς εκεί φενετε καλυτερα τα αποτελέσματα καθε κανονικης εκφρασης.

## Περιγραφή των Κανονικών Εκφράσεων

1. '<title>([^(</title>)]+)</title>' στην αρχη ψαχνω το ταγκ title και μετα προσθετω ολα μεχρι το κλεισιμο του ταγκ title, ετσι διαλέγω και ότι βρίσκεται ανάμεσα στα ταγκ. Τελος, διαλέγω και ταγκ κλεισιματος.
2. '<!--([^->]+)-->' πρώτα ψαχνω να έχει το ανοιγμα σχολείων μετά διαλέγω τα παντα μεχρι να βρω το κλείσιμο των σχολείων και τέλος διαλέγω και το κλείσιμο των σχολείων.
3. '(<(script)|(style)[^>]\*>)[^<]\*(</(script)|(style))>' στην αρχη ψαχνω ένα ταγκ script ή style μετα διαλέγω τα παντα μεχρι τον χαρακτηρα '<' που θα είναι ο πρώτος χαρακτηρας του ταγκ κλεισιματος(ο λογος που εβαλα μονο αυτο τον χαρακτηρα είναι επειδη προσπαθησα να κανω τον κωδικα να αναγνωρισει το κλεισιμο ταγκ για τα script, style αλλα ποτε δεν μου εβγαζε το επιθυμητο αποτελεσμα). Τελος, διαλέγω το ταγκ κλεισιματος για τα script η style.
4. '<a .\* (href="[^\"]\*" )' στην αρχη διαλέγω το <a και ότι είναι αναμεσα του και του href. Μετα διαλέγω το href= και συνεχιζω να διαλέγω ότι ακολουθεί μεχρι πριν βρώ το συμβολο ". Στη συνέχεια διαλέγω και το συμβολο ". Από το href μεχρι το

5. '<[>]+>)' αρχικά ψαχνω το σύμβολο "<" μετά διαλέγω τα πάντα μέχρι πριν βρω το ">". Τέλος, διαλέγω και το ">".
6. '&((amp)|(gt)|(lt)|(nbsp));' στην αρχή ψαχνω το "&". Μετά, ψαχνω να δω αν υπάρχει amp ή gt ή lt ή nbsp. Το κάθε ένα από τα προηγούμενα είναι σε παρενθέσεις για να κάνει την σύγκριση σε όλη την συμβολοσειρά και όχι μόνο σε ένα χαρακτήρα. Όλα τα πράγματα που είναι σε με ή είναι όλα μαζί σε μια παρένθεση ώστε μεσο του group(1) να καταλαβαίνω ποιο από αυτά έχει βρεθεί στο αποτέλεσμα της κανονικής έκφρασης και να το αντικαταστήσω με τον κατάλληλο χαρακτήρα. Τέλος, υπάρχει και ένα ",".
7. '\s+' μου βρίσκει όλα τα 1+ συνεχόμενα κενά.

Οι μονές πηγες που χρησιμοποιηθηκαν ήταν τα λινκ στο github που δοθηκαν με την εκφώνηση της εργασίας