Περιγραφή υλοποίησης κάθε ερωτήματος ξεχωριστά:

Αρχικά ανοίγω το txt αρχείο και με τη χρήση file descriptor αποθηκεύω σε μεταβλητή τα περιεχόμενά του.

1)Κανονική έκφραση: <title>(.+?)</title>

Βρίσκει οτιδήποτε είναι ανάμεσα στο <title> και στο </title>.

Έχω χρησιμοποιήσει παρένθεση στο .+? έτσι ώστε να τυπώσω το περιεχόμενο με το group(1).

2)Κανονική έκφραση:<!.+?>

Βρίσκει οτιδήποτε είναι ανάμεσα στο <! >.

Στη συνέχεια πραγματοποιώ απαλοιφή με τη χρήση της sub().

3)Κανονική έκφραση: (<script.+?</script>)|(<style.+?</style>)

Βρίσκει οτιδήποτε είναι ανάμεσα στα <script </script> ή στα <style </style>.

Στη συνέχεια πραγματοποιώ απαλοιφή με τη χρήση της sub().

4)Κανονική έκφραση: <a(.+?)

Βρίσκει οτιδήποτε ανάμεσα στα <a και με τη χρήση της παρένθεσης εξάγω τα περιεχόμενα με το group(1).

5)Κανονική έκφραση:<.+?>

Βρίσκει οτιδήποτε ανάμεσα στα < > δηλαδή το περιεχόμενο των tags .

Στη συνέχεια πραγματοποιώ απαλοιφή με τη χρήση της sub().

6)Κανονική έκφραση: (&|>|<|)

Με τη χρήση της εναλλαγής αναγνωρίζει ένα από αυτά κάθε φορά.

Έχω φτιάξει ένα dictionary όπου τα keys είναι τα HTML entities και τα values είναι τα σύμβολα που θέλουμε να τα αντικαταστήσουμε.

Στη συνέχεια φτιάχνω μία συνάρτηση callback όπου με την if/elif ταυτίζω το κάθε HTML entity με το group(1) της έκφρασης και στη συνέχεια επιστρέφω το αντίστοιχο value από το λεξικό.

Η αντικατάσταση επιτυγχάνεται με την χρήση της sub().

7)Κανονική έκφραση:\s+

Αναγνωρίζει ένα και παραπάνω συνεχόμενους κενούς χαρακτήρες.

Στη συνέχεια πραγματοποιώ απαλοιφή με τη χρήση της sub().