

Μεταγλωτιστές 2020

Προγραμματιστική Εργασία 2

Όνοματεπώνυμο: Χαράλαμπος Θεοδούλου
ΑΜ: Π2017151

Πρώτα έβαλα το re της python. Στη συνέχεια έφτιαξα τις μηχανές ταιριάσματος, τις κανονικές εκφράσεις και μια συνάρτηση για τη αντικατάσταση `& > < `. Τέλος διαβάζουμε το `testpage.txt`, τρέχουμε τις κανονικές εκφράσεις και τη συνάρτηση και τα τυπώνουμε.

Ερωτημα 1:

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

Με την `.` Και με `+` γίνεται το ταίριασμα των χαρακτήρων που είναι ανάμεσα στο `<title></title>`. Η παρένθεση είναι για να τα κρατήσουμε.

Ερωτημα 2:

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

Γίναντε απαλοιφή των σχολίων. Μέσο του `group(0)` πιάνουμε ολόκληρα τα σχόλια.

Ερώτημα 3:

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

Μέσο `.` και `*` ταιριάζουμε αυτά που είναι μέσα στο `script` και `style`. Επίσης χρησιμοποιώ το `/1` του `group(1)`.

Ερωτημα 4:

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

Ταίριασμα περιεχομένου `href` και `a`.

Ερώτημα 5:

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

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

Ταίριασμα `<></>`.

Ερωτημα 6:

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

Μέσο αυτής είναι τα `& > < `.

Ερώτημα 7:

```
(r'\s+')
```

Ταίριασμα των `whitespace`.