

## Εργασία 2 – Report

Ονοματεπώνυμο: ΠΑΠΑΠΕΤΡΟΥ ΑΘΑΝΑΣΗΣ

ΑΜ: Π2017149

1ο Ερώτημα: ('<title>(.\*?)</title>')

Η χρήση της . και + χρησιμοποιηθήκαν για να ταιριάζουν ότι είναι ανάμεσα στο <title> και το </title>.

2ο Ερώτημα: ('<!--.\*?-->',re.DOTALL)

Η χρήση της . και \* χρησιμοποιηθήκαν για να ταιριάζουν ότι είναι ανάμεσα στο <!-- και το -->. Δεν χρειάζεται να έχουμε τουλάχιστον ένα χαρακτήρα γιατί βάλαμε το \* και όχι το +.

3ο Ερώτημα: (r'<(s(?:cript|tyle)).\*?>.\*?</\1>',re.DOTALL)

Η χρήση της <(s(?:cript|tyle) χρησιμοποιηθήκαν για να ταιριάζουν τα tags <script></script> και <style></style>.

4ο Ερώτημα: (r'<a.+?href="(.\*?)".\*?>(.\*?)</a>',re.DOTALL)

Η χρήση της (r'<a.+?href="(.\*?)".\*?>(.\*?)</a>',re.DOTALL) χρησιμοποιήθηκε για να ταιριάξει την href και αυτά που περιέχει μέσα η <a></a>.

5ο Ερώτημα: (r'<.+?>|</.+?>',re.DOTALL) και (r'<.+?/>',re.DOTALL)

Η χρήση των πιο πάνω εντολών είναι να βρίσκει τι γράφει μέσα στις <>.

6ο Ερώτημα: (r'&(amp|gt|lt|nbsp);')

Η χρήση της | είναι για να ταιριάξει όλα όσα είναι μέσα στην παρένθεση amp, gt ,lt ,nbsp.

7ο Ερώτημα: (r'\s+')

Η χρήση του + είναι για να συνδυάσουμε ένα ή περισσότερους χαρακτήρες.