

Ερώτημα 1ο : ('<title>(.*?)</title>')

Η χρήση της . Και του + χρησιμοποιούνται για να ταιριάσουν ο,τι είναι μεταξύ του <title> και </title>.

Ερώτημα 2ο : ('<!--.*?-->',re.DOTALL)

Η χρήση της . Και του * χρησιμοποιούνται για να ταιριάσουν ο,τι είναι ανάμεσα σε <!-- & → Δεν χρησιμοποιούμε παρενθέσεις γιατί απαιτεί την απαλοιφή ολόκληρου του σχόλιου.

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

Χρησιμοποιείτε το (s(?:cript|tyle) για να ταιριάζουν τα tags <script></script> και <style></style>.

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

Αποτελείτε το ταίριασμα της ιδιότητας href και το περιεχόμενο των <a> .

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

Ερώτημα 6ο : (r'&(amp|gt|lt|nbsp);')

Χρησιμοποιούμε το | για να ταιριάζουν κάθε φορά μια από τις τέσσερις επιλογές που έχουμε πιο πάνω.

Ερώτημα 7ο : (r'\s+')

Είναι για να ταιριάζουν συνεχόμενοι χαρακτήρες , με το + είναι για να ταιριάζουν ένας η περισσότεροι χαρακτήρες whitespace.