

Μεταγλωτιστές 2020
Προγραμματιστική Εργασία #2

Ονοματεπώνυμο: Χαράλαμπος Σωκράτης Βαφάκος
ΑΜ: Π2017015

- 1) Εξαγωγή και εκτύπωση του τίτλου (οτιδήποτε βρίσκεται μεταξύ <title> και </title>)
 - a) <title>(.+?)</title> | Κανονική έκφραση
 - b) Βρίσκει ότι είναι μέσα στο <title> και </title>
- 2) Απαλοιφή και εκτύπωση του τίτλου (οτιδήποτε βρίσκεται μεταξύ<>)
 - a) <!.+?> | Κανονική έκφραση
 - b) Βρίσκει ότι είναι μέσα στο <! >
- 3) Απαλοιφή των <scripts> και <style> tags με όλο τους το περιεχόμενο
 - a) (<scripts.+?</scripts>)|(<style.+?</style>) | Κανονική έκφραση
 - b) Βρίσκει ότι είναι μέσα στα <scripts></scripts> και <style></style>
- 4) Εξαγωγή και εκτύπωση του συνδέσμου από <a> tags και του κειμένου τους
 - a) <a(.+?) | Κανονική έκφραση
 - b) Βρίσκει ότι είναι μέσα στα <a
- 5) Απαλοιφή όλων των tags από το κείμενο
 - a) <.+?> | Κανονική έκφραση
 - b) Βρίσκει ότι είναι μέσα στα <>
- 6) Μετατροπή των ειδικών HTML entities
 - a) (&|>|<| |"|'|©|®) | Κανονική έκφραση
 - b) Κάθε φορά αναγνωρίζει ένα από τα παραπάνω
- 7) Μετατροπή ακολουθιών συνεχόμενων χαρακτήρων whitespace σε κενό
 - a) \s+ | Κανονική έκφραση
- 8) Print