ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ

Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΑΝΑΠΤΥΞΗ ΜΗΧΑΝΙΣΜΩΝ ΑΥΤΟΜΑΤΗΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΝΟΜΙΚΩΝ ΚΕΙΜΕΝΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Μάριος Κόνιαρης, 210 772 1402, mkoniari@dblab.ece.ntua.gr
Γιώργος Παπαστεφανάτος, 210 6875403, gpapas@imis.athena-innovation.gr

ΠΕΡΙΛΗΨΗ: Σκοπός της εργασίας είναι η μελέτη και ανάπτυξη μηχανισμών αυτόματης κατηγοριοποίησης σε κείμενα νομικής φύσης.

ΑΤΟΜΑ: 1
ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα τελευταία χρόνια η αλματώδης ανάπτυξη της πληροφορικής έχει διευρύνει σε σημαντικό βαθμό τον όγκο και τη διακίνηση πληροφορίας, η οποία καθίσταται προσβάσιμη σε ολοένα και μεγαλύτερο αριθμό χρηστών. Μέσα σε αυτόν τον κυκεώνα διακινούμενων σε ηλεκτρονική μορφή δεδομένων, ο χρήστης συχνά δυσκολεύεται να εντοπίσει και να αντλήσει την πληροφορία που τον ενδιαφέρει. Αναδεικνύεται επομένως το αίτημα ανάπτυξης κατάλληλων τεχνικών που θα διευκολύνουν την προσπέλαση και τη διαχείριση της διαθέσιμης πληροφορίας ανάλογα με τις ανάγκες των χρηστών. Το αίτημα αυτό επιχειρεί να ικανοποιήσει και ο κλάδος της Αυτόματης Κατηγοριοποίησης Κειμένου (Automated Text Categorization), δηλαδή της αυτόματης κατάταξης κειμένων γραμμένων σε φυσική γλώσσα σε ένα σύνολο προκαθορισμένων κατηγοριών.

Το αντικείμενο της διπλωματικής είναι η υλοποίηση ενός συστήματος που θα επιτρέπει την αυτόματη κατηγοριοποίηση νομικών κειμένων (Νομοθεσία / Νομολογία).

Στα πλαίσια της διπλωματικής εργασίας θα πραγματοποιηθούν οι ακόλουθες εργασίες:

- Θα μελετηθεί η σχετική βιβλιογραφία και θα συζητηθούν προσεγγίσεις/ιδέες για την αυτόματη κατηγοριοποίηση νομικών κειμένων.
- Θα εξεταστεί η δυνατότητα χρήσης του EuroVoc (πολύγλωσσος θησαυρός της Ευρωπαϊκής Ένωσης)
- Θα υλοποιηθούν αλγόριθμοι και κριτήρια κατηγοριοποίησης κειμένων και θα αξιολογηθεί η αποτελεσματικότητά τους.

Για την υλοποίηση της Δ.Ε., οι υποψήφιοι θα έχουν στην διάθεση τους την Ελληνική νομοθεσία σε xml μορφή.

Η συγκεκριμένη εργασία θα πραγματοποιηθεί σε συνεργασία με το Ινστιτούτου Πληροφοριακών Συστημάτων του Ερευνητικού Κέντρου "Αθηνά".

ΣΧΕΤΙΚΟ ΥΛΙΚΟ

[1]. C.C. Aggarwal and C.X. Zhai. A survey of text classification algorithms. Mining Text Data, pages 163–222, 2012. http://charuaggarwal.net/text-content.pdf

[2]. Efficient pairwise multi-label classification for large-scale problems in the legal domain. In Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.160.6028&rep=rep1&type=pdf)

[3]. Eurovoc (http://eurovoc.europa.eu/drupal/?q=el)