

**ΑΝΑΠΤΥΞΗ ΜΗΧΑΝΙΣΜΩΝ ΑΥΤΟΜΑΤΗΣ ΕΞΑΓΩΓΗΣ ΚΑΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ
ΑΝΑΦΟΡΩΝ ΣΕ ΝΟΜΙΚΑ ΚΕΙΜΕΝΑ (AUTOMATIC REFERENCE RESOLUTION IN
LEGAL TEXTS)**

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 εξαγωγής, κατηγοριοποίησης και ανάκτησης συσχετίσεων με βάση αναφορές σε κείμενα νομικής φύσης

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η αναφορά είναι μια διευκρίνιση που επιτρέπει τον εντοπισμό ενός παραθέματος, τη μνεία ενός έργου ή μέρους έργου που έχει σχέση με το αντικείμενο του κειμένου ή την παραπομπή, μέσα στο ίδιο κείμενο ή σε υποσημειώσεις. Οι αναφορές που απαντώνται στα νομικά κείμενα συνήθως ακολουθούν μια αυστηρή δομή, γεγονός που επιτρέπει την αυτόματη εξαγωγή τους βασιζόμενοι σε ένα σύνολο κανόνων [1].

Το αντικείμενο της διπλωματικής είναι η υλοποίηση ενός υποσυστήματος που θα επιτρέπει την αυτόματη συσχέτιση κειμένων με βάση τις αναφορές που συναντά. Το υποσύστημα αυτό θα αποτελέσει επέκταση υπάρχοντος συστήματος οργάνωσης, διαχείρισης και ανάκτησης της Ελληνικής νομοθεσίας με τεχνολογίες του Σημασιολογικού Ιστού. Δεδομένα εισόδου για το υποσύστημα θα είναι κείμενα νομικής φύσεως, σε δομημένη xml μορφή. Το υποσύστημα θα πρέπει να ανιχνεύει σε αυτά τις αναφορές που απαντώνται, όπως αναφορές στην νομοθεσία της Ελλάδος/ Ευρωπαϊκής Ένωσης, αναφορές σε Δικαστικές αποφάσεις, βιβλιογραφικές αναφορές κτλ. Εν συνεχεία οι αναφορές αυτές θα κατηγοριοποιούνται κατάλληλα και θα εντοπίζεται το αναφερόμενο κείμενο. Το αρχικό κείμενο/ αναφερόμενο κείμενο θα τροποποιούνται καταλλήλως για να συμπεριλάβουν τις ανιχνευμένες αναφορές (πχ προσθήκη υπερσυνδέσμου με anchor html tag)

Τα στάδια εκπόνησης της διπλωματικής εργασίας προβλέπονται ως εξής:

- Διερεύνηση και Κατηγοριοποίηση των Αναφορών.
- Εντοπισμός και Εξαγωγή Αναφορών
- Ανάκτηση Αναφορών
- Πειραματική Επαλήθευση αποτελεσμάτων (Δείκτες precision, recall)

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 Ινστιτούτο Πληροφοριακών Συστημάτων του Ερευνητικού Κέντρου "Αθηνά".

ΣΧΕΤΙΚΟ ΥΛΙΚΟ

[1]. Γενική Γραμματεία Της Κυβέρνησης - Εγχειρίδιο οδηγιών για την κωδικοποίηση της νομοθεσίας Σελ 12 (http://www.ggk.gov.gr/wp-content/uploads/2010/02/teliko_egxeiridio_odigion_gia_tin_kodikopoiisi_tis_nomothesias.pdf)

[2]. M. Martínez-González, "Reference extraction and resolution for legal texts," PReMI'05 Proc. First Int. Conf. Pattern Recognit. Mach. Intell., pp. 218–221, 2005.
http://link.springer.com/chapter/10.1007/11590316_29