



Ανάλυση Δεδομένων Ληματογράφησης (Σημασιολογική Διασύνδεση)

Περιγραφή

Σκοπός της εργασίας είναι να αξιοποιήσει ένα σύνολο δεδομένων τα οποία έχουν προκύψει από την ψηφιακή ληματογράφηση ενός λεξικού ώστε να αναλυθεί ο τρόπος με τον οποίο κάθε λήμμα μπορεί να διασυνδεθεί σημασιολογικά με άλλα λήμματα σύμφωνα με την ανάλυση του κειμένου του!

Η διαδικασία ληματογράφησης περιλαμβάνει ένα σύνολο από μεταδεδομένα μεταξύ των οποίων ο ορισμός και η ανάπτυξη. Τα δύο αυτά στοιχεία είναι «κειμενικά» δηλαδή αποτελούνται από καθαρό κείμενο.

Σκοπός της εργασίας είναι να αναπτυχθεί ένα λογισμικό το οποίο θα παίρνει σαν είσοδο σύνολα κειμένων λημάτων και θα εφαρμόζει αλγορίθμους συσχέτισης κειμένων (ML) με σκοπό να εντοπίσει σημασιολογικές συσχετίσεις μεταξύ των κειμένων. Μπορούν να χρησιμοποιηθούν παραδοσιακές μετρικές ομοιότητας ή και τεχνικές βασισμένες σε word embeddings.

Το λογισμικό που θα αναπτυχθεί θα έχει τη δυνατότητα να παίρνει σαν είσοδο ένα αρχείο που περιλαμβάνει τα μεταδεδομένα των λημάτων και να εξάγει λίστες με τις συσχετίσεις μεταξύ τους (σύμφωνα με τη χρήση κάθε αλγορίθμου που θα επιλεγεί).

Παραδοτέα

- Το λογισμικό σε πλήρη λειτουργία
- Ο κώδικας του λογισμικού
- Η αναφορά της πτυχιακής εργασίας.

Απαραίτητες και επιθυμητές γνώσεις

Καλή γνώση προγραμματισμού (python, java).

Πλήθος φοιτητών

1-2 άτομα.

Επιβλέπων καθηγητής: Αναπληρωτής Καθηγητής Βασίλειος Πουλόπουλος (vacilos@uop.gr)

Προτείνεται η επαφή με τον επιβλέποντα πριν δηλώσετε το παρόν θέμα ώστε να σας είναι ξεκάθαρη η ακριβής έκταση της εργασίας και οι απαιτήσεις της.

Η μέγιστη διάρκεια για την ολοκλήρωση της πτυχιακής εργασίας είναι ένα ημερολογιακό έτος.