



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ: **Αναγνώριση Προφίλ Συγγραφέα με χρήση Τεχνικών Επεξεργασίας Φυσικής Γλώσσας (Author Profiling Using Natural Language Processing Techniques)**

ΕΠΙΒΛΕΠΩΝ ΔΙΔΑΣΚΩΝ: **Εν. ΣΤΑΜΑΤΑΤΟΣ**

ΠΕΡΙΛΗΨΗ

Στόχος Διπλωματικής Εργασίας:

Η εργασία αυτή στοχεύει στην μελέτη τεχνικών επεξεργασίας φυσικής γλώσσας και την εφαρμογή τους στην αναγνώριση χαρακτηριστικών ενός συγγραφέα όπως το φύλο και η ηλικία.

Περιγραφή Διπλωματικής Εργασίας:

Σε πολλές εφαρμογές που σχετίζονται με κείμενα που αναρτώνται σε κοινωνικά δίκτυα είναι πολύ χρήσιμο να εξάγουμε πληροφορία σχετική με το φύλο, την ηλικία ή άλλα δημογραφικά χαρακτηριστικά του συγγραφέα. Αυτή η εργασία καλείται αναγνώριση προφίλ συγγραφέα (author profiling) και αντιμετωπίζεται συνήθως με χρήση τεχνικών αναπαράστασης κειμένου που βασίζονται σε χαμηλού επιπέδου πληροφορία (συχνότητες λέξεων ή ακολουθιών χαρακτήρων). Η εργασία αυτή αποσκοπεί στην εξέταση χαρακτηριστικών υψηλότερου επιπέδου που μπορεί να προκύψουν από τεχνικές επεξεργασίας φυσικής γλώσσας (natural language processing) όπως η συντακτική και η σημασιολογική ανάλυση των κειμένων. Θα χρησιμοποιηθούν υπάρχουσες βιβλιοθήκες μεθόδων επεξεργασίας φυσικής γλώσσας και μηχανικής μάθησης και θα πραγματοποιηθεί πειραματική μελέτη για την αξιολόγηση της μεθόδου.

Σκελετός της εργασίας:

- Μελέτη τεχνικών author profiling
- Μελέτη τεχνικών επεξεργασίας φυσικής γλώσσας
- Σχεδίασης και ανάπτυξη προτεινόμενης μεθόδου
- Πειραματική αξιολόγηση της μεθόδου

Απαιτούμενες Γνώσεις:

Καλή γνώση προγραμματισμού (Python) και αλγορίθμων μηχανικής μάθησης.

Βιβλιογραφία:

- Stamatatos, E. A Survey of Modern Authorship Attribution Methods, Journal of the American Society for Information Science and Technology, 60(3), pp. 538-556, 2009, Wiley.
- A. Pastor López-Monroy, M. Montes-y-Gómez, H.J. Escalante, L. Villaseñor-Pineda, and E. Stamatatos, Discriminative Subprofile-Specific Representations for Author Profiling in Social Media, Knowledge-Based Systems, 89, pp. 134–147, Elsevier, 2015.