

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Ευστάθιος Σταματάτος

Α. Επαλήθευση Συγγραφέα με Χρήση Προ-εκπαιδευμένων Γλωσσικών Μοντέλων

Συνοπτική περιγραφή:

Η επαλήθευση συγγραφέα (authorship verification) είναι ένα βασικό πρόβλημα στο χώρο της ανάλυσης/αναγνώριση συγγραφέα όπου ο στόχος είναι να αποφασίσουμε αν δύο κείμενα έχουν γραφεί από τον ίδιο συγγραφέα ή όχι¹. Η εργασία θα εστιάσει στη χρήση προεκπαιδευμένων γλωσσικών μοντέλων (pre-trained language models)², π.χ. BERT, και στην ανάλυση σύντομων εδαφίων (προτάσεις). Ιδιαίτερη έμφαση θα δοθεί σε περιπτώσεις δια-τομεακής επαλήθευσης (cross-domain verification) όπου τα δύο κείμενα έχουν διαφορετικά χαρακτηριστικά (π.χ. email vs. essay, tweet vs. blog). Μέσω πειραμάτων σε υπάρχουσες συλλογές κειμένων θα διερευνηθεί κατά πόσο μπορεί να γίνει βελτίωση της επίδοσης χρησιμοποιώντας με κατάλληλο τρόπο τα διαθέσιμα προεκπαιδευμένα γλωσσικά μοντέλα.

¹ <https://doi.org/10.1002/asi.21001>

² <http://dx.doi.org/10.18653/v1/N19-1423>