

Εκτίμηση κι Επικύρωση Εγκυρότητας Περιεχομένου, με τη Χρήση Μεθόδων/Αλγορίθμων Μηχανικής Μάθησης

Επιστημονικές Περιοχές:

Κύρια Περιοχή: Επικύρωση εγκυρότητας περιεχόμενου (content verification) & Μηχανική μάθηση

Δευτερεύουσα Περιοχή: Εντοπισμός ψευδών ειδήσεων (fake news) και παραπληροφόρηση

Περιγραφή:

Στο πλαίσιο της διπλωματικής αυτής εργασίας θα γίνει σχεδιασμός κι ανάπτυξη ενός συστήματος εκτίμησης και επικύρωσης διαδικτυακού περιεχομένου (online content verification) με τη χρήση αλγορίθμων μηχανικής μάθησης. Τα δεδομένα που θα εξεταστούν θα αφορούν κυρίως διαδικτυακό περιεχόμενο όπως ειδήσεις, αναρτήσεις σε μέσα κοινωνικής δικτύωσης, κ.α. τα οποία θα είναι διαθέσιμα είτε μέσω ελεύθερα διαθέσιμων συνόλων δεδομένων (datasets) είτε θα ανακτώνται μέσω κάποιου API. Σκοπός ανάπτυξης αυτού του συστήματος θα είναι η ανάλυση και η αξιολόγηση διαδικτυακού περιεχομένου ούτως ώστε να εντοπιστεί ψευδές και παραπλανητικό περιεχόμενο εφαρμόζοντας σύγχρονες μεθόδους μηχανικής μάθησης, στο πλαίσιο έρευνας και ανάπτυξης στον τομέα της καταπολέμησης της παραπληροφόρησης και των ψευδών ειδήσεων.

Η πορεία υλοποίησης της διπλωματικής θα περιλαμβάνει τα κάτωθι στάδια:

- Μελέτη και αναζήτηση της σχετικής με το θέμα της διπλωματικής βιβλιογραφίας για τον εντοπισμό και τη μελέτη των state-of-the-art εργαλείων και μεθόδων.
- Ανάπτυξη κώδικα για την ανάκτηση/εισαγωγή δεδομένων διαδικτυακού περιεχομένου που βρίσκονται διαθέσιμα είτε διαδικτυακά (publicly available) είτε θα ανακτηθούν από αντίστοιχα APIs.
- Προ-επεξεργασία ανακτημένων δεδομένων και μετασχηματισμός αυτών (όπως εξαγωγή χαρακτηριστικών, κ.α.) σε μορφή που εξυπηρετεί τις ανάγκες των αλγορίθμων ανάλυσής τους
- Αναζήτηση και δοκιμή μοντέλων επικύρωσης και επαλήθευσης περιεχομένου
- Σύγκριση και αξιολόγηση αποτελεσμάτων για την επιλογή της αποδοτικότερης μεθόδου για την αξιολόγηση και επαλήθευση διαδικτυακού περιεχομένου.

Απαιτήσεις:

- Γνώση προγραμματισμού (Python, SQL, NoSQL)
- Εξοικείωση με απλές έννοιες ανάλυσης περιεχομένου κι εφαρμογές τύπου REST
- Τήρηση χρονοδιαγράμματος-Υψηλή διαθεσιμότητα

Αριθμός ατόμων:

1

Στοιχεία επικοινωνίας:

N. Πεππές (npeppes@cn.ntua.gr), Θ. Αλεξιάκης (talexakis@cn.ntua.gr), Ε. Δ. Συκάς (sykas@cn.ntua.gr)