

## Εξαγωγή Σημασιολογικού Λεξικού για Ανάλυση Συναισθήματος

### Περιγραφή

Τα συστήματα Ανάλυσης Συναισθήματος Βασισμένης σε Χαρακτηριστικά (Aspect Based Sentiment Analysis - ABSA) δέχονται ως είσοδο ένα σύνολο κειμένων (π.χ. κριτικές προϊόντων ή μηνύματα κοινωνικών δικτύων) που σχολιάζουν μια συγκεκριμένη οντότητα (π.χ. ένα συγκεκριμένο μοντέλου κινητού τηλεφώνου) και επιχειρούν να εντοπίσουν τα κυριότερα χαρακτηριστικά (aspects) της οντότητας που γίνονται αντικείμενο σχολιασμού (π.χ. μπαταρία, οθόνη), καθώς και να εκτιμήσουν το μέσο συναίσθημα (π.χ. πόσο θετικό ή αρνητικό είναι) που εκφράζουν τα κείμενα για κάθε χαρακτηριστικό της οντότητας.

Στην ανάπτυξη τέτοιων συστημάτων χρησιμοποιούνται συχνά Λεξικά Συναισθηματος, σύνολα δηλαδή όρων και λέξεων συνδεδεμένα με μία ή περισσότερες τιμες που ποσοτικοποιούν την πολικότητα του συναισθηματος που εκφραζουν (π.χ. "νοστιμος" -> θετικό, "αγενής" -> αρνητικό). Συχνά όμως ένας ορος μπορεί να εχει διαφορετική πολικότητα ανάλογα με την οντότητα και τα χαρακτηριστικά το οποια αξιολογει. Π.χ. ο όρος "υψηλός" εχει θετικη χροιά οταν χαρακτηριζει το την ικανότητα ενος μαγειρα αλλά αρνητική οταν χαρακτηριζει την πίεση ενός ασθενή. Τα υπάρχοντα λεξικά δεν κανουν εκ των προτέρων αυτή τη διακριση, αναθέτοντας στους ορους που περιεχουν μια μοναδική τιμή πολικότητας.

Στο πλαίσιο αυτό, στόχος της συγκεκριμένη διπλωματικής εργασίας είναι η αναπτυξη ενός συστηματος για την (ημι-αυτοματη) κατασκευή σημασιολογικών λεξικών συναισθήματος τα οποία, δεδομένου ενός συγκεκριμένου γνωστικού πεδίου, περιεχουν την εξής πληροφορία:

- Τις οντότητες του πεδίου και τα χαρακτηριστικά αυτών για τα οποία μπορεί να εκφραστεί συναίσθημα (π.χ. η ποιότητα του φαγητού για εστιατόρια ή η διάρκεια μπαταρίας για laptops ).
- Τους όρους που μπορούν να χρησιμοποιηθούν για την εκφραση συναισθήματος για καθένα από αυτά τα χαρακτηριστικά (π.χ. "νοστιμο" για το φαγητό).
- Την πολικότητα (θετική ή αρνητική) που έχουν τα συγκεκριμένα ζεύγη χαρακτηριστικών-όρων.

Η εργασία θα καλύψει τουλάχιστον 2 γνωστικά πεδία (εστιατόρια και laptops) και ανάλογα με την ποιότητα των αποτελεσμάτων μπορεί να οδηγήσει σε επιστημονική δημοσίευση. Η μέγιστη διάρκεια ολοκλήρωσης της πτυχιακής είναι ένα ημερολογιακό έτος.

### Παραδοτέα

- Σύστημα που λειτουργεί
- Πηγαίος κώδικας με αναλυτική τεκμηρίωση
- Αναφορά πτυχιακής εργασίας (σε LaTeX)

### Απαραίτητες και επιθυμητές γνώσεις

Προγραμματισμός σε Java και εξοικείωση με αλγορίθμους και εργαλεία επεξεργασίας φυσικής γλώσσας ή/και μηχανικής μάθησης είναι απαραίτητα για την ολοκλήρωση της συγκεκριμένης εργασίας.

### Βιβλιογραφία

- [1] P. Alexopoulos, M. Wallace (2015), "Creating Domain-Specific Semantic Lexicons for Aspect-Based Sentiment Analysis", 10th International Workshop on Semantic Media and Adaptation (SMAP), November 5-6, 2015, Trento, Italy
- [2] M Pontiki, D Galanis, H Papageogiou, S Manandhar, I Androutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, Proceedings of the 9th International Workshop on Semantic Evaluation

### Πλήθος φοιτητών

1 ή 2 άτομα. Η ακριβής έκταση της εργασίας θα είναι ανάλογη του αριθμού των φοιτητών που θα την αναλάβουν.

### Επιβλέποντες

Πάνος Αλεξόπουλος ([palexopoulos@expertsystem.com](mailto:palexopoulos@expertsystem.com)) και Εμμανουήλ Γουάλλες ([wallace@uop.gr](mailto:wallace@uop.gr)). Προτείνεται η επαφή με τους επιβλέποντες πριν δηλώσετε το παρόν θέμα, ώστε να σας είναι ξεκάθαρη η ακριβής έκταση της εργασίας και οι απαιτήσεις της.