

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Σγόντζος Κωνσταντίνος
Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Αναπλ. Καθηγητής, Ι. Τζιτζικας**

**Παρασκευή, 15/02/2019, 10:00
Αίθουσα Β108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**“ Υψηλής Ποιότητας Ανάκτηση Σχολίων Χρηστών στα Πλαίσια της Διαλογικής
Εξερευνητικής Αναζήτησης”**

ΠΕΡΙΛΗΨΗ

Τα Συστήματα Απάντησης Ερωτήσεων (Question Answering Systems) μπορούν να θεωρηθούν ως επέκταση των Μηχανών Αναζήτησης, υπό την έννοια ότι προσπαθούν να απαντήσουν τις ερωτήσεις των χρηστών που τίθενται σε φυσική γλώσσα, αντί απλά να επιστρέφουν πόρους συναφείς με τις λέξεις της ερώτησης. Η χρήση τέτοιων συστημάτων αυξάνεται ραγδαία τα τελευταία χρόνια και έχουν πολλά πεδία εφαρμογής (e-commerce, και αλλού).

Το γενικό πλαίσιο αυτής της εργασίας είναι η στήριξη της διαδικασίας *Εξερευνητικής Αναζήτησης* στη μορφή διαλόγου σε φυσική προφορική γλώσσα. Συγκεκριμένα επικεντρωνόμαστε στο πλαίσιο της *Πολυεδρικής Αναζήτησης (Faceted Search)*, η οποία είναι μία μορφή *εξερευνητικής αναζήτησης* που συχνά εφαρμόζεται σε πολυδιάστατα δεδομένα ή δεδομένα γραφικής δομής. Ωστόσο τυχαίνει μια ερώτηση του χρήστη να μην

μπορεί να απαντηθεί από τα δεδομένα, και σε αυτήν την περίπτωση τυχόν αδόμητη πληροφορία σε μορφή κειμένου, όπως σχόλια και κριτικές από χρήστες, αποτελούν μια πολύτιμη πηγή πληροφοριών. Στην εργασία αυτή, επικεντρωνόμαστε στο πως θα χρησιμοποιήσουμε αυτά τα έγγραφα (κριτικές και σχόλια χρηστών) για την απάντηση των ερωτήσεων του χρήστη. Τα έγγραφα αυτά περιορίζονται συνήθως στο μέγεθος μιας πρότασης ή μιας παραγράφου και είναι ευρέως γνωστά ως Παράγραφοι Απαντήσεων (*Answer Passages*). Λόγω της αλληλεπίδρασης με το χρήστη με προφορικό λόγο, απαιτείται υψηλής ποιότητας ανάκτηση σχολίων.

Στην εργασία αυτή προτείνουμε διάφορες μεθόδους για τον εντοπισμό των πιο συναφών σχολίων χρηστών. Για τη μελέτη του προβλήματος φτιάξαμε μια συλλογή αξιολόγησης, και προσαρμόσαμε μια υπάρχουσα του TREC στις ανάγκες του προβλήματός μας. Εν συνεχεία προτείναμε τρία βασικά μοντέλα, και παραλλαγές αυτών. Το μοντέλα βασίζονται σε ευρέως διαδεδομένα χειρόγραφα (Wordnet) και στατιστικά (Word2vec) λεξικά. Τα συγκριτικά πειραματικά αποτελέσματα έδειξαν ότι: (α) ο συνδυασμός χρήσης χειρόγραφων και στατιστικών λεξικών επιτυγχάνει μεγαλύτερα ποσοστά ακρίβειας στο πρόβλημα της εύρεσης συναφούς παραγράφου, (β) ο εμπλουτισμός της ερώτησης του χρήστη με περαιτέρω σχετικούς όρους, σε συνδυασμό με ανακατανομή των βαρών των όρων της, βασισμένη σε λέξεις που περιγράφουν το γενικό πλαίσιο του συνόλου των πληροφοριών, αυξάνουν την απόδοση του συστήματος, κυρίως σε όταν έχουμε ένα συγκεκριμένο πεδίο, όπως αυτό της κράτησης ξενοδοχείου. Σε σχέση με τις αντίστοιχες εργασίες, οι μέθοδοι που προτείνουμε επιτυγχάνουν υψηλότερο βαθμό ακρίβειας.

Konstantinos Sgontzos

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Associate Professor, I. Tzitzikas

Friday 15/02/2019, 10:00

Room B108, Computer Science Dept., University of Crete

“On Achieving High Quality User Reviews Retrieval in the Context of Conversational Faceted Search”

ABSTRACT

Question Answering Systems can be considered as an extension of Search Engines in the sense that they directly answer questions posed by users in Natural Language instead of simply returning relevant sources based on keyword search. Recently, such systems gained more attention and have great potential in various application domains including e-commerce.

The general context of our work is to support *Faceted Search (FS)* in the form of spoken dialogue. *FS* is an exploratory search paradigm that is commonly applied over multidimensional or graph data. However, since structured data might not suffice for answering a user’s query, unstructured information like user comments or reviews can be a valuable source of information. In this work we focus on how to leverage the textual information (user reviews and comments) for answering the user questions. This information is generally restricted in the length of a sentence or a paragraph and is known as *Answer Passage (AP)*. Due to the speech interaction with the user, high quality user review retrieval is necessary.

To study this problem, we created an evaluation collection and we adjusted to our problem an existing TREC collection. We then proposed three basic approaches and variations of them. These approaches are based on handmade (Wordnet) and statistical (Word2vec) dictionaries. The comparative experimental results show that: (a) The combination of handmade and statistical dictionaries succeed better results in terms of precision in relevant *AP* (review) retrieval, (b) The expansion of the question with other relevant terms combined with methods that devalue question words which describe the general domain context, improve the precision, mainly in restricted domains such as *hotel booking*. In comparison to existing related work, our proposed approaches achieve higher precision.