

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Δημητράκης Ελευθέριος
Μεταπτυχιακός Φοιτητής**

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

Τρίτη, 19/02/2019, 15:00

Αίθουσα Β108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

**“Απάντηση Ερωτήσεων Ανοιχτού Πεδίου πάνω από εκατοντάδες Ανοιχτά
Συνδεδεμένα Σύνολα Δεδομένων”**

ΠΕΡΙΛΗΨΗ

Η Απάντηση Ερωτήσεων Ανοιχτού Πεδίου (Open Domain Question Answering) αποτελεί πρόκληση, και απαιτεί μεταξύ άλλων, την αντιμετώπιση του προβλήματος της κατανομής των δεδομένων, δηλαδή του γεγονότος ότι τα σύνολα δεδομένων είναι διάσπαρτα σε πολλά διαφορετικά μέρη. Στην εργασία αυτή, εστιάζουμε στην Απάντηση Ερωτήσεων Ανοιχτού Πεδίου (Open Domain Question Answering), αξιοποιώντας Ανοιχτά Συνδεδεμένα Δεδομένα (Linked Open Data). Περιοριζόμαστε σε τρία (3) είδη ερωτήσεων: ερωτήσεις γεγονότων, ερωτήσεις επιβεβαίωσης και ερωτήσεις ορισμού. Παρουσιάζουμε και αξιολογούμε συγκριτικά, διαδικασίες που βασίζονται στην Εξαγωγή Πληροφοριών (Information Extraction) με σκοπό την απάντηση ερωτήσεων. Ιδιαίτερο χαρακτηριστικό της προσέγγισής μας είναι η δυνατότητα απάντησης επερωτήσεων για εκατομμύρια οντότητες, εκμεταλλευόμενοι ταυτόχρονα εκατοντάδες πηγές Συνδεδεμένων Δεδομένων (Linked Data), χωρίς να απαιτείται η χρήση δεδομένων για εκπαίδευση (training data). Η μέθοδος αποτελείται από τρεις κύριες φάσεις: (α) Ανάλυση Ερώτησης (η οποία περιλαμβάνει την αναγνώριση του τύπου της ερώτησης καθώς και τον καθαρισμό της), (β) Εντοπισμός Οντοτήτων, Αναγνώριση Ονομασίας Οντοτήτων και τη Σύνδεση

αυτών με τις υποκείμενες πηγές, (γ) Εξαγωγή Απάντησης (η οποία περιλαμβάνει την ανάκτηση RDF τριπλετών, τη βαθμολόγηση αυτών και την εξαγωγή της καλύτερης τριπλέτας). Επιδεικνύουμε τα οφέλη της προσέγγισης αυτής όσον αφορά το πλήθος των ερωτήσεων που μπορούν να απαντηθούν καθώς και την επαλήθευση των απαντήσεων. Επίσης ερευνούμε, μέσω πειραματικών αποτελεσμάτων, πως επηρεάζουν τα βήματα της Απάντησης Ερωτήσεων (Question Answering) την αποτελεσματικότητα. Η αξιολόγηση βασίστηκε σε 1000 ερωτήσεις από το SimpleQuestions σύνολο δεδομένων καθώς και 2500 ερωτήσεις από το QALD-7 Large-Scale σύνολο δεδομένων.

Lefteris Dimitrakis

M.Sc. Thesis

Computer Science Department

University of Crete

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

Tuesday 19/02/2019, 15:00

Room B108, Computer Science Dept., University of Crete

“Open Domain Question Answering over Hundreds of Linked Open Datasets”

ABSTRACT

Open domain Question Answering is a challenging task that requires, among others, to tackle the data distribution issue, i.e. the fact that datasets are scattered in several places. In this thesis, we focus on open domain Question Answering over Linked Open Data. We confine ourselves to three kinds of questions: factoid, confirmation, and definition questions. We introduce and comparatively evaluate information extraction based processes for question answering. The distinctive feature of our approach is that it can answer questions over millions of entities, by exploiting hundreds of Linked Data sources simultaneously, without having to use any training data. The process comprises three main phases:

(i) Question Analysis (that includes question type identification and cleaning), (ii) Entities Detection, Named Entity recognition and linking and (iii) Answer Extraction (that includes RDF triples retrieval, scoring and matching). We demonstrate the benefits of this approach in terms of answerable questions and answer verification, and we investigate, through experimental results, how the steps of the question answering process affect the effectiveness of question answering. The evaluation was based on 1000 questions from SimpleQuestions and 2500 from QALD-7 Large-scale datasets.