

Title: Question and Answer Classifier for Closed Domain Interactive Question Answering

Author: Đinh Lê Thành

Supervisors: Dr. Raffaella Bernardi, RNDr. Markéta Lopatková, Ph.D.

## **ABSTRACT**

Nowadays natural language processing has made big progress thanks to the application of statistical approaches and to the large amount of data available to train the systems. These progresses are pushed by the several evaluation campaigns. Thanks to them systems are compared and progress measured. These evaluations are mostly based on data sets artificially developed by the organizers of such evaluation campaigns. In our work we show that though useful these data sets are biased and there is the need of developing data generated in a more natural setting by real users. We consider as case studies the classification of questions. In particular we look at the classification of questions types needed in Question Answering systems, and the classification of follow up questions into topic continuation and topic shift needed in Interactive Question Answering. We evaluate classifiers first on TREC data and than on a corpus of real user's data. In both cases the performance of the classifiers drops significantly showing the need of working on more users centered systems. The results also show that the classifiers could be better fine tuned taking into account the new challenges real users data launch to NLP systems. We leave this for future research.

Keywords: taxonomy, question classification, topic shift, topic continuation, follow-up.