

# A Relevant Passage Retrieval and Re-ranking Approach for Open-Domain Question Answering

Nouha Othman\*, Rim Faiz\*\*

\*Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia  
othmannouha@gmail.com

\*\*Université de Carthage IHEC Carthage, LARODEC, Tunisia  
rim.faiz@ihec.rnu.tn

**Abstract.** Question answering systems (QAS)s aim to directly return precise answers to natural language questions. Retrieving and re-ranking passages are viewed as the most challenging tasks in a typical QAS and still require non-trivial effort. In this paper, we propose a novel approach for retrieving and re-ranking passages using n-grams and SVM. Our n-gram based passage retrieval engine relies on a new measure of similarity between a passage and a question. The retrieved passages are further re-ranked using a Ranking SVM model combining different text similarity measures in order to return the most relevant passage to a given question. Our experiments and results have shown promise and demonstrated that our approach is competitive.

## 1 Introduction

Over the past few years, with the ongoing development in information technology, the amount of data has been increasing massively every day. Thereby, recent years have witnessed a burgeoning interest in question answering (QA) which is one of the major research area under Information Retrieval (IR) with applications mainly from Information Extraction (IE) and Natural Language Processing (NLP) (Faiz, 2006), where the main aim is to directly deliver a precise and concise answer to a question posted by the user in natural language from a sizable collection of documents or database (Voorhees, 2001). Indeed, QA domain falls into two categories: closed-domain QA which deals with questions under a specific domain such as biology and medicine, and open-domain QA which deals with general questions in various domains without any limitation. In our work, we focus on the second category as the techniques used are not tailored toward a specific domain. Basically, a typical QAS can be broadly viewed as a pipeline which consists of four main modules (Tellex et al., 2003): question analysis, document retrieval, passage retrieval and answer extraction, where each module has to deal with specific challenges. In particular, passage retrieval (PR) is always considered as the key task of a typical QAS since it allows to reduce the search space from a large collection of documents to a fixed number of passages. Obviously, QASs cannot find the right response to a given question, unless the answer exists in one of the retrieved passages. Therefore, it was widely studied over the past few years. Re-ranking passages is also a crucial task at the end of the QAS pipeline which aspires to order the retrieved passages such that the most relevant