

Abstract

The semantic web contains a large amount of data in the form of knowledge graphs (KGs). KGs, made in Resource Description Framework (RDF) data model, provide a way to represent knowledge from heterogeneous, non-integrated, and inconsistent data. Simple Protocol and RDF Query Language (SPARQL), a query language, is used to build queries and retrieve the information to access the data in KGs. Working with SPARQL to express information needs and explore KGs is difficult for users who are unfamiliar with query languages and the structure of the underlying KGs. Consequently, question answering over KGs has emerged to make them more accessible and seek answers to questions expressed in a natural language manner.

This thesis aims to delve deep into the research of answering natural language questions posed by users across KGs. Question answering over KGs is a challenging task that targets designing systems that are capable of answering users' questions using the RDF facts stored in KGs. This thesis focuses on providing approaches to retrieve answers to factoid questions, either simple or complex.

To assist end-users in accessing KGs created on small and medium scales, a semantic parsing-based approach is proposed, including the offline phase and the semantic parsing phase. The main objective of this approach is to generate training samples in the offline phase and then use the samples to train machine learning models employed in the semantic parsing phase to parse a given input question and automatically find the corresponding SPARQL query. To showcase the proposed approach, a KG describing touristic entities is applied. Furthermore, a graph-driven approach is proposed to bridge the gap between natural language questions and SPARQL queries in huge KGs. Given a question, a graph is built to represent the question's intention as well as a subgraph of the underlying KG is extracted to reduce the huge search space. Then, the question graph is matched over the subgraph to find a query graph corresponding to a SPARQL query, and finally, the query is executed to return the answers. Experiments are conducted to evaluate the approach and exhibit its results.

The achieved results of the proposed approaches indicate the performance in terms of recall, precision, and F1-score and improvements compared to the state-of-the-art. However, there still exists much space for improvement of question answering over KGs.