Linguistic rule-based translation of natural language question into SPARQL query for effective semantic question answering

ABSTRACT

Semantic question answering (SQA) demands different processing compared to the common information retrieval method because the semantic knowledge base is stored in the triples form. However, manipulating the knowledge requires understanding of its structure and proficiency in semantic query language such as SPARQL. Natural language interface (NLI) alleviates this by allowing user to input question in their human language. Then it produces an answer by translating the input into an equivalent SPARQL before it is executed to retrieve the answer. However, none of the existing research has presented a holistic computational model for the translation of NL question into an equivalent SPARQL for the semantic KB querying. Existing studies have focused mainly on the semantic disambiguation through consolidation where user interaction is initiated so that relevant concept can be chosen by the user to be inserted into the SPARQL. Besides, the linguistic understanding of the input has limited coverage where only one triple is constructed which loses many original expressions. There is a necessity to increase the linguistic understanding to involve multi-variables and multi-triples in the translated SPARQL so that accurate answer will be returned. Therefore, in this paper the linguistic challenge in NLI is addressed, specifically on the question complexity depth, processes that need to be performed to answer the question and gaps in existing study. A linguistic-rule-based translation model for natural language question is introduced that utilizes a set of observational variables to extract the information in the question; (i) checking if the focus is equals to subject, (ii) number of subjects, (iii) number of property, (iv) number of object, (v) checking if object is instance, (vi) checking if the question contains superlative expression, (vii) superlative orientation and (viii) checking if the question contains aggregates expression. The model is also aimed to reduce dependability on clarification dialogues. The results show that the approach has managed to eliminate clarification dialogues and increase linguistic coverage of NLI.

Keyword: Natural language; Natural language interface; Semantic question answering; Semantic search; Semantic web; SPARQL