SWFQA Semantic Web Based Framework for Question Answering

Irphan Ali, Uttarakhand Technical University, Dehradun, India
Divakar Yadav, Madan Mohan Malviya University of Technology, Gorakhpur, India
Ashok Kumar Sharma, BSA Institute of Technology and Management, Faridabad, India

ABSTRACT
A question answering system aims to provide the correct and quick answer to users’ query from a knowledge base. Due to the growth of digital information on the web, information retrieval system is the need of the day. Most recent question answering systems consult knowledge bases to answer a question, after parsing and transforming natural language queries to knowledge base-executable forms. In this article, the authors propose a semantic web-based approach for question answering system that uses natural language processing for analysis and understanding the user query. It employs a “Total Answer Relevance Score” to find the relevance of each answer returned by the system. The results obtained thereof are quite promising. The real-time performance of the system has been evaluated on the answers, extracted from the knowledge base.

KEYWORDS
Knowledge Base (KB), Named Entity Recognizer (NER), Natural Language Processing (NLP), Query Processing, Question Answering (QA), Question Classification, Semantic Web (SW)

1. INTRODUCTION
The existing QA systems answer basic queries. Techniques for answering user’s complex queries are being developed. The existing techniques analyze and parse complex queries into basic queries and utilize existing methods for answering them (Abacha, 2015). However, finding the correct answer is amongst the most critical issues in QA frameworks. Recent works demonstrate that increasing the performance of QA framework is subject to the quality of likely answers present in the repository. In fact, this gap has prompted us to associate the user query with related context to find an appropriate answer in real time.

The semantic web (SW) has been assumed as next step to World Wide Web (Tim Berners-Lee, 2001). Semantic web technology utilizes an arrangement of universal standards as presented by global community, the World Wide Web Consortium (W3C). The semantic web technology builds on the following main standards such as Resource Description Framework (RDF), SPARQL protocol and Web Ontology Language (OWL). RDF is a format used by semantic web technology to store information on the semantic web. SPARQL is an RDF query language able to extract and manipulate data stored in RDF formats and OWL is a computational logic-based language with the end goal that information presented in OWL can be utilized by PC programs.

Though, the proposed framework is designed for an open domain, but we have taken dataset from E-learning domain to carry out the experiments. E-learning questions are viewed as more complex questions as they require domain information and furthermore long answers need to be retrieved from

DOI: 10.4018/IJIRR.2019010106

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
various documents. These questions have inborn ambiguity. The presence of unlimited documents on the Web, a combined approach to natural language processing (NLP) techniques and information retrieval (IR) has gained fame for the development of question answering frameworks.

The solution to the problem specified above is to use a dedicated QA framework that answers user’s natural language questions. This is difficult for questions that are bound to a domain and the domain under consideration is unrestricted and no assumption can be sensibly made (Ajitkumar, 2016).

The proposed framework addresses the problem of answering English questions, formulated in natural language. There are various types of questions, but we only considered questions expressed by wh- pronouns such as who, what, where, when, how much and how many. An answer can be an E-learning entity for these questions. Moreover, proposed framework associates a context with a question and thereafter provides an appropriate answer. It also does user assisted answer validation to improve the rank of answers for future interactions. The rest of the paper is organized as follows: related work in section 2, the proposed QA framework in Section 3, experimental evaluation in section 4 and discussion & comparative review in section 5. We finally concluded the work in Section 6.

2. RELATED WORK

BASEBALL (Green, 1961) and LUNAR (Woods, 1973) is among the first known question answering systems (QAS). It could answer queries regarding dates, locations, and American baseball games. LUNAR was one of the primary scientific QAS. It helped the geographical examination of the stones brought by the Apollo mission and accurately addressed 90% of the queries posed by users.

The presence of semantic information in the web pages permits machines to process such information and helps users to search, share and merge the information more conveniently (Shen, 2018). Natural language-based question answering systems for semantic web generates relations between words in document repository as well as finds the accurate answers (Saint-Dizier & Moens, 2011; Melo, 2018; Subalalitha, 2017). Moreover, this system uses the knowledge of reasoning to interpret the returned document, to get the correct answer corresponding to user’s query without his/her help.

The LASSO QAS (Moldovan, 1999) answers open domain queries by returning short and long answers. It incorporates three modules: question processing (finding query type, expected answers, query focal point, and query keyword), paragraph indexing and answer processing. A query type classification was developed manually from the examination of the TREC-8 data over 25 query types. LASSO utilized 8 requested heuristics to retrieve and identify named entities such as people, associations, areas, dates, monetary standards, and items. To assess the correctness of each user answer, seven scores are calculated. The appropriate answer is extracted by picking the appropriate user answer with the most appropriate combined score.

The PRECISE framework examines natural language queries and converts them into SQL queries (Popescu, 2003). It refers to an arrangement of semantic question classes and converts user queries that belong to at least one of the semantic queries to output the related SQL queries. In its assessment, PRECISE gave accurate answers to more than 80% of the queries in the domains of jobs, geography, and restaurants. PRECISE is effective on semantic queries which present 77.5% in the geography database and 97% in restaurant database.

The START framework (Katz, 2002) answers queries regarding geography with an accuracy of 67% on 3,26,000 questions. START utilizes knowledge bases, enhanced manually to identify triplets of the form (subject, relation, object). It extracts answers based on the interrelation of the user query with the annotations of the knowledge bases.

AquaLog (Lopez and Motta, 2004) is an ontology-based QAS which enables users to pick ontology as input and to ask queries in natural language. It returns answers from the accessible semantic markup by joining some techniques together such as to understand natural language questions regarding the domain covered by the ontology and to outline them to semantic markup. AquaLog is a convenient
Semantic Text Summarization Based on Syntactic Patterns

[www.igi-global.com/article/semantic-text-summarization-based-on-syntactic-patterns/109660?camid=4v1a](www.igi-global.com/article/semantic-text-summarization-based-on-syntactic-patterns/109660?camid=4v1a)