Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 117 (2017) 183–191

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Arabic Computational Linguistics, ACLing 2017, 5-6 November 2017, Dubai, United Arab Emirates

Arabic Question Answering Using Ontology

Ali Albarghothi[a], Feras Khater[a], Khaled Shaalan[a] *

[a]*Faculty of Engineering & IT, The British University in Dubai, Dubai, United Arab Emirates*

Abstract

Currently, with the massive amount of data that is being posted on the web at a rapid pace, users regularly have inquiries and they expect to find out short and precise answers. Semantic web and ontology technologies are becoming the essential components to represent domain-specific data that can be utilized in Question Answering System (QAS). In this paper, we introduce an Arabic QAS based on the domain knowledge or ontology in order to answer natural language inquiries. Prior to the implementation, it was crucial to perform some Natural Language Processing (NLP) tasks that assisted in analyzing the questions such as normalization, tokenizing, removing the stop words, stemming and tagging. Furthermore, we present how to develop the ontology through the Protégé tool, how to translate the inquiries into triple patterns and build the SPARQL queries which are the mechanism to retrieve the answer from Resource Description Framework (RDF) data. As far as Arabic is concerned, Arabic Question Answering (QA) is still limited and did not reach the similar level of English QA because of challenges on the Arabic language, for example, complexity in morphology derivational and inflectional, and words suffer from the scare of vowels. In addition, Arabic language orthography does not use capital letters or the like, which affect on Named Entity Recognitions. Hence, it is an opportunity to focus on Arabic QA using ontology toward getting a clear concept of this semantic-based approach. The result of the experimental results demonstrates the feasibility of constructing a QAS based on ontology. The proposed model has achieved promising results with accuracy of 81%, which provides an important indication for further in-depth study and analysis.

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Arabic Computational Linguistics.

Keywords: Arabic Question Answering; Questions Answering System; Natural Language Processing; Semantic web; Ontology; Resource Description Framework; SPARQL.

* Corresponding author. Tel.: +971502551461 fax: +0-000-000-0000 .
E-mail address: alibarg31@yahoo.com, feraskhater@gmail.com, Khaled.shaalan@buid.ac.ae

1877-0509
Peer-review under responsibility of the scientific committee of the 3rd International Conference on Arabic Computational Linguistics.
10.1016/j.procs.2017.10.108

1. Introduction

Nowadays, the amount of information published on the Internet is so huge, users regularly have inquiries and expect to find out the short and precise answer. Users generally express the inquiries in their own language without limiting to a particular query language or formation rules. QAS provides the capability to answer the user question which is posted in natural language. NLP is a part of linguistic and Artificial Intelligence (AI), which focuses on the interaction between computer machine and people based on human language [2].

With the revolution of Internet, users discovered it is harder to get their fundamental data on the web rapidly and successfully, which has become a vital issue. Designing a special query or adding the semantics to web resources is the best ideal approach to enhance the quality of data search. Semantic Web is the best approach to present the semantics of documents and empower these semantics to be used by web applications [5]. Semantic Web search attempts to increase and enhance conventional search results by using ontologies that assume a major part in this sort of search. By definition, ontology is a formal model of the regular understanding of the entities and relations of interest [6]. Fensel [3] mentioned that Sir John Berners-Lee proposed the seven layers structure of Semantic Web (UNICODE & URI, XML, Ontology, RDF, Proof, Logic, Trust). XML, RDF, and ontology are considered as the backbone layers that are utilized to describe the semantics of web data.

Ontology is an official representation and explicit description of concepts in a certain knowledge domain and the relation between those concepts. The basic elements of ontologies contain (individuals, relations, classes, attributes, restrictions, rules, function terms, axioms, and events) which describe the real life and play vital roles in data processing [10].

Applying QAs in Arabic language finds many challenges and difficulties, these challenges refer to language structure [7]. Inflectional and derivational in Arabic language that causes shortened in Arabic statements. Additionally, diacritics in Arabic language makes an ambiguity to QA especially in Question Analysis and Answer Analysis methods. Conversely, the Latin languages are easier that do not have diacritics in their orthography. Also, Arabic language does not use capital letters which effect on Named Entity Recognitions more difficult [9].

There are a numerous number of researches that have been conducted in English QA systems. Furthermore, other studies have been conducted in other languages (Spanish, Spain, etc.). As per our research, AbuTaha & Alagha [1] designed a prototype model for Arabic QA using semantic web which enables the user to query using NL and RDF backends. From this point, we found some improvement areas which motivate us to develop Arabic QA using ontology system handling linguistic and semantic processing.

AbuTaha & Alagha [1] developed a prototype for Arabic QA using ontology by translating pre-defined NL queries into SPARQL using simple rules. This ontology built on Pathology domain to be as a knowledge repository contains a sample of some diseases information. Furthermore, the ontology used contains the relationship between the main classes, objects and data properties. Also, their model utilizes four components that are Question processing, Data processing, Answer retrieval and Ontology mapping. Ontology rules were defined to identify the dependencies between all ontology objects and to support answer retrieval. They performed an experimental test on 30 questions to translate Arabic NL queries into SPARQL. Also, the author analyzed all pre-defined questions manually to build the SPARQL queries. The 30 queries executed on Jena platform and achieved 28 correct answers. The author suggests implementing a hybrid system handling multiple domains with increasing the number of questions, rules and type of questions (factoid and complex questions), as well as developing a combination of linguistics and semantic processing.

In this paper, we introduce QAS based on the domain knowledge or ontology in order to answer natural language inquiries. Furthermore, we present how to develop the ontology through the Protégé tool, how to translate the inquiries into triple patterns and build the SPARQL queries which are the mechanism to retrieve the answer from Resource Description Framework (RDF) data. The proposed model has achieved promising results, which provides an important indication for further in-depth study and analysis.

2. Methodology

The methodology that we have followed in this study can be divided into four stages. First, building the dataset using semantic web and ontology by using Protégé tool in order to generate the ontology dictionary as RDF/OWL file [8]. Second, the questions are selected based on the ontology information, the number of questions is 100 varying