

The Role of ICT in our Daily Life Applications: Obstacles and Challenges

A Holy Quran Ontology Construction with Semi-Automatic Population

Eman Elkhammash and Wahiba Ben Abdessalem

Department of Computer Science, Faculty of Computer and Information Technology, University of Taif, Taif, Saudi Arabia

ABSTRACT

Ontology population is an instantiation of the ontology classes and subclasses. Ontology population is the main step in ontology construction. However, the manual population is a time-consuming task. Accordingly, automatic or semi-automatic methods to populate an ontology are required. This paper suggests an approach for the creation of an ontology and its population. The studied ontology is related to named entities in the holy Quran. The major contribution of this approach is to harness the benefits of learning methods, conjoined with statistical models to extract contexts (words surrounding a named entity) from Quran and Hadith and retain the weighty contexts for the recognition of supplementary named entities to populate the ontology.

KEY WORDS: ONTOLOGY; HOLY QURAN; NAMED ENTITY; MACHINE LEARNING

INTRODUCTION

The Holy Quran is a document having its proper and unique style. It is a base knowledge discussing practically all life fields. Ontology is a type of knowledge representation and it became a major device for numerous applications that are interested in the semantic content. A holy Quran ontology will offer a powerful representation of its knowledge. An ontology with named entities

of the holy Quran will be useful for implementing many applications (information retrieval, information extraction). The population of this ontology is also a huge task.

This paper aims to create an ontology related to named entities in the Holy Quran (names of god, angels, prophets...) from the holy Quran and prophetic traditions (Hadith). Hadith is considered as a very important source of Islamic knowledge. In our work, we preferred to use Sahih Al-Bukhari [4] and Sahih Muslim [5] collections to

ARTICLE INFORMATION:

*Corresponding Author: Eman.kms@tu.edu.sa,
wahiba.bak@gmail.com

Received 12th Nov, 2018

Accepted after revision 29th Dec, 2018

BBRC Print ISSN: 0974-6455

Online ISSN: 2321-4007 CODEN: USA BBRCBA

 Thomson Reuters ISI ESC / Clarivate Analytics USA and
Crossref Indexed Journal

NAAS Journal Score 2018: 4.31 SJIF 2017: 4.196

© A Society of Science and Nature Publication, Bhopal India
2018. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

DOI: 10.21786/bbrc/12.1/26

extract instances of the concepts described by the ontology.

This work proposes, also, a new approach, which is Natural Language Processing (NLP)-independent process for a semi-automatic population of a Quran ontology from texts (Quran and Hadith). This approach applies learning techniques and statistical models to acquire and classify ontology instances.

The article is organized into five sections as follows. Section 2 summarizes related work. Section 3 presents the proposed approach related to holy Quran ontology construction and population. Section 4 describes the implementation experiments. Finally, section 5 concludes the work.

RELATED WORK

Several studies have been undertaken on the topic of Quran ontology. In this section, we introduce some researches that describe Quranic ontology. Quran corpus ontology that describes around 300 concepts in Quran and 350 relations between them. It is available at (<http://corpus.quran.com/ontology>). Quran corpus focuses only on concepts mentioned in Quran and provides features such as Arabic grammar, syntax, and morphology for each word in the Holy Quran [2].

Another work in [1] developed an OWL ontological model for Quranic concepts described in the Quran. SPARQL queries are then used to retrieve the knowledge from the Ontology where single queries can extract similar concepts of the holy Quran which are spread out over different chapters and verses. There are numerous researches that developed a particular chapter in Quran or focused on particular subjects such as [3] that described domain ontology of living creatures and introduced semantic search of the living creatures mentioned in the Holy Quran including animals and birds. Also [9] proposed an ontology for "Salat (prayers)" based on translated texts of the Quran.

All the above mentioned works either focused on particular subjects in Quran or built ontology based on several concepts existing in Quran, but did not use approaches or methods to ensure that the conceptualisation is comprehensive.

The ontology developed in this paper augments concepts from Quran and Hadith. Moreover, we provide in this paper a new approach to further extend the developed ontology using learning methods.

PROPOSED APPROACH

A large number of methods exist for ontological engineering, describing the steps for the construction of an ontology. The building of the Holy Quran ontology was

directed by the University of Stanford [6]. It comprises the following steps:

- Step1: Definition of the ontology domain and scope.
- Step2: Exploration of the possibility to reuse existing ontologies.
- Step 3: Identification of the ontology terms.
- Step 4: Definition of the ontology hierarchy of classes.
- Step 5: Definition of the classes' attributes.
- Step 6: Definition of the attributes' values.
- Step 7: Population of the ontology.

The steps we follow to develop the proposed holy Quran ontology baptized: NEQ-Ontology are as follows.

Step 1: Definition of the ontology domain and scope: this step allows determining the ontology domain by answering (A) a set of basic questions (Q), such as:

Q: What is the domain that the ontology will cover?

A: Named entities in the Holy Quran and Hadith.

Q: For what we will use the ontology?

A: For many applications dealing with the holy Quran, such as information retrieval, information extraction, text mining...

Q: What types of questions will the ontology answers to?

A: The names of God, angels, devil, hell, paradise...

Q: Who will use and maintain the ontology?

A: Researchers interested in the study of the holy Quran and Hadith

Step 2: Exploration the possibility to reuse existing ontologies: the ontology "corpus. Quran" [7] is showing a classification of 300 concepts that appear in the Quran. This ontology is used to retrieve semantically relevant verses. However, this ontology does not hold all the class' instances. For example, the god names are missing, only the name "Allah" exists. Also, some angels 'names are indicated in the authentic texts: Sahih ElBukhari and Sahih Muslim (i.e. Israfeel, Monkir, Nakeer) do not exist in this ontology.

We will reuse some of the concepts that appear in this ontology [7], rather than develop a new one from scratch. However, we will make a new taxonomy and populate this ontology based on the Holy Quran and Hadith of Sahih ElBukhari and Sahih Muslim.

Step 3: Identification of the ontology terms: a complete list of the terms (in a non-structured form) related to the ontology domain is defined. For example God, angels, prophet, religion, language, Holy Book, food, disease, animal, body organ, mosque, vegetable, tree, Planets, events, month, day, location, mountain, color, etc.

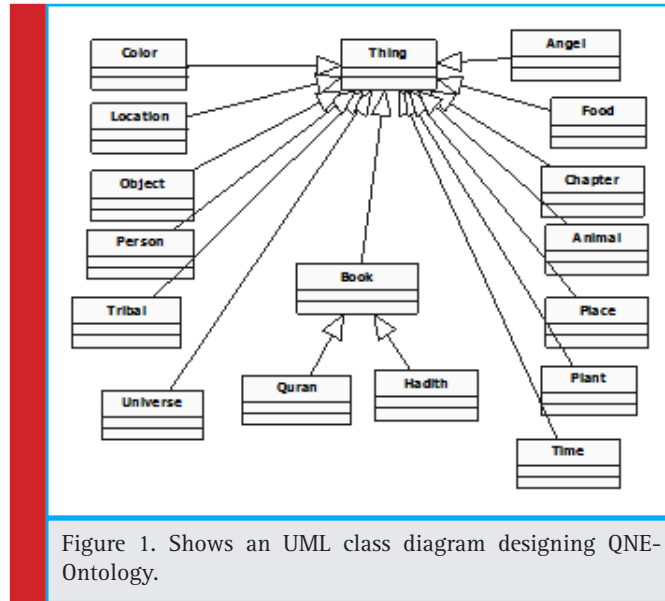


Figure 1. Shows an UML class diagram designing QNE-Ontology.

Step 4: Definition of the ontology hierarchy of classes: in this step, based on the terms defined in the previous step, classes are defined in a hierarchical taxonomy.

Step 5: Definition of the classes' attributes: The properties are attached to classes or subclasses taking into account the inheritance between classes and subclasses.

Step 6: Definition of the attributes' values: The value types of the attributes are assigned to attributes. In addition, the multiplicity (the number of values for attributes) is determined, and the relations between classes.

Step 7: Population of the ontology: in order to create the classes' instances of the ontology, we used a semi-automatic method. The process for the semi-automatic ontology population is proposed in the following algorithm:

1. $Instances \leftarrow Examples$
2. $List-Occ \leftarrow Find-occ(Instances, Quran, Sahih El Bukhari, Sahih Muslim)$
3. $Cont \leftarrow Find-Context(Instances, Quran, Sahih El Bukhari, Sahih Muslim)$
4. $RankC \leftarrow Rank-Context(Cont, Quran, Sahih El Bukhari, Sahih Muslim, Instances)$
5. $Instances \leftarrow Find-Exp(Cont, Quran, Sahih El Bukhari, Sahih Muslim)$
6. $Instances-Validation(instances)$
7. *Return to step 2.*

Our approach for populating the Holy Quran ontology with named entities performs as follows: (1) we introduce for each concept a small set of confident and well-known instances as examples of named entities (i.e.

Jebreel, Mikaeel). Then, (2) we search all occurrences of those instances on the Holy Quran and Hadiths. For these occurrences, (3) we identify words surrounding them. We call context of a named entities, the set of n words that precede and follow the named entity. (4) We calculate the ranking of contexts to assure the quality of these contexts. The aim of this step is to identify the most pertinent contexts that appear in the texts. The ranking function is computed as follows:

$$w_i = c_{fi} * ic_{fi} * \sum e_{fi} * \sum ie_{fi}$$

Where, (c_{fi}) is the context frequency, (ic_{fi}) is the inverse context frequency, (e_{fi}) is the example frequency, and (ie_{fi}) the example inverse frequency. These frequencies are computed based on the formula: Term Frequency-Inverse Document Frequency (*tf-idf*) [8]. (5) Search segments matching the retained contexts. These segments are considered as newly named entity instances. (6) The found named entity instances are validated by an expert in Islamic knowledge and the initial set of instances is updated and extended by larger examples: the validated instances. (7) Perform a new iteration. In each iteration, the system learns, generates news contexts, and therefore new instances.

IMPLEMENTATION

The Holy Quran ontology (NEQ-Ontology) is implemented using the ontology editor Protégé 2000 [9]. The ontology consists of 15 super classes (Figure 2):

Individuals are used to represent the objects (instances) of the classes. For instance, *Al_Quran*, *Bible*, *Torah* are represented as individuals in the Class *Book*.

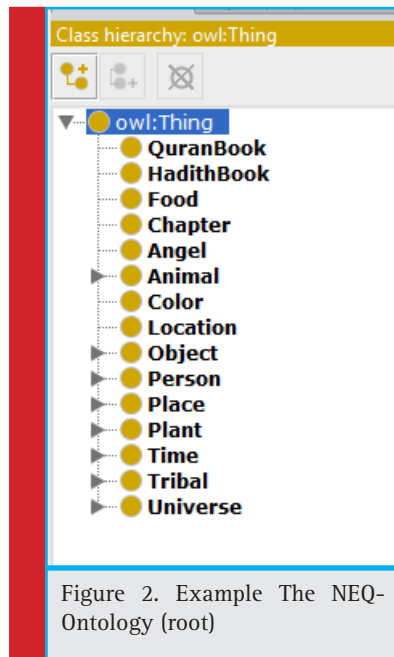


Figure 2. Example The NEQ-Ontology (root)



Figure 3. Sample of the individuals.

Mohammed, Musa, Essa are individuals in the subclass Prophet under Person class. Al-Qaswa is an individual of Cow subclass under Animal class. Figure 3 shows some individuals of NEQ ontology.

Properties are binary relations on individuals. Figure 4 shows sample of object properties of NEQ ontology.

For instance, in the context of the story mentioned in the holy Quran that describes the Almighty order to Bani Israel to slaughter a cow. We have *Miraculous Cow*,

which is an individual in the COW subclass under the Animal class, and it is linked to *Yellow* individual via the object property *has Color* to describe the color of the *Miraculous Cow*. The individual *Musa* under *Prophet* class is linked to the individual *Bani Israel* of *Tribal* class via the property *SenderTo*. Moreover, the individual

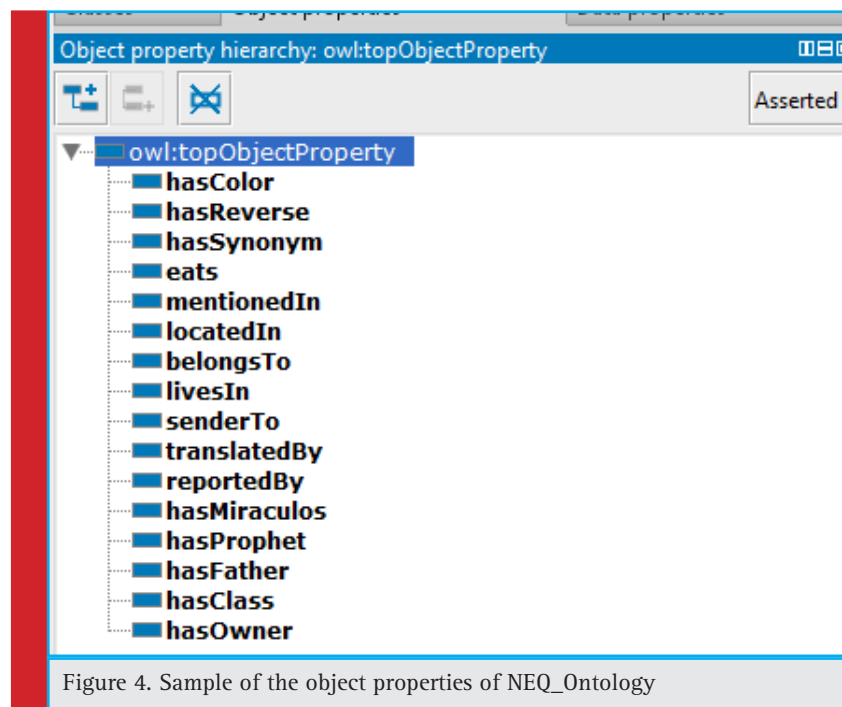


Figure 4. Sample of the object properties of NEQ_Ontology

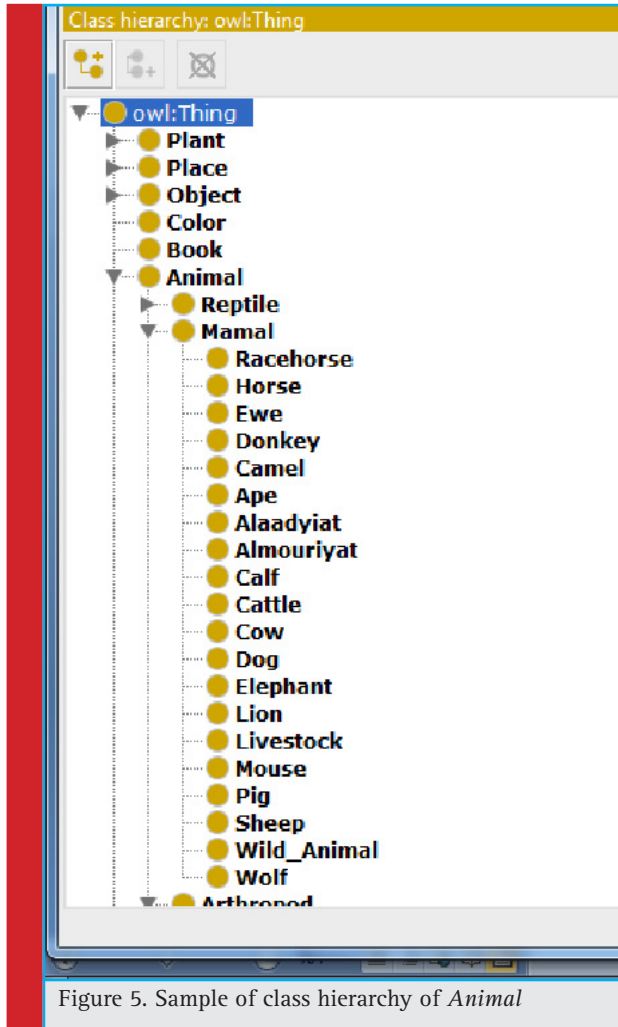


Figure 5. Sample of class hierarchy of *Animal*

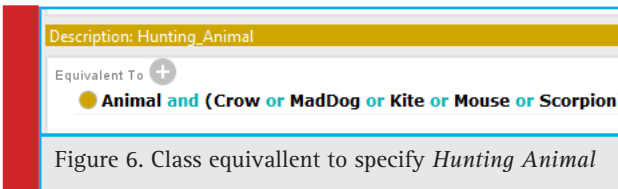


Figure 6. Class equivalent to specify *Hunting Animal*

Musa is linked to the individual *Miraculous Cow* via the property *has Mircaulous*.

A class can extend another class, for instance, *Animal* class in Figure 5 can be divided into a number of subclasses such as, *Reptile*, *Mammal*, *Arthropod*... subclasses and *Mammal* class can further be extended into a number of subclasses such as *Horse*, *Ewe*, etc.



Figure 7. OWL for Class equivalent of Figure 4

According to Sahih Muslim *Hunting animal* are five which is *Crow* or *Mad Dog* or *Kite* or *Mouse* or *Scorpion*. We use class equivalent, which is a built-in property that links a class description to another class description to define *Hunting animal* as shown in Figure 6.

The OWL formalism of corresponds to the definition of *HuntingAnimal* can be shown in Figure 7.

Protégé offers a number of visualization tools that facilitate visualization of the ontology structure. Figure 8 depicts a view of *Camel* class, which is a subclass of *Animal* and has an individual call *Al_Qaswa*.

We used OntoGraf tool to visualize the ontology. OntoGraf is a visualization tool in Protégé which provides many features to visualize the structure of the whole ontology or part of it. It also supports various layouts to organize the structure of the ontology and view different relationships. For instance, the tooltip in Figure 9 showed extensive details about the properties of individual *Al_Qaswa* such as the source that mentioned *Al-Qaswa*, *Al_Qaswa* meaning,...etc.

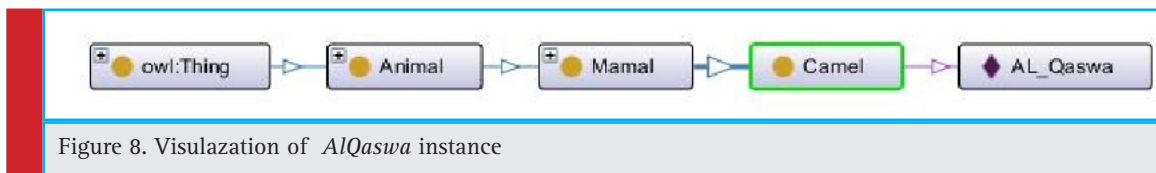


Figure 8. Visulazation of *AlQaswa* instance

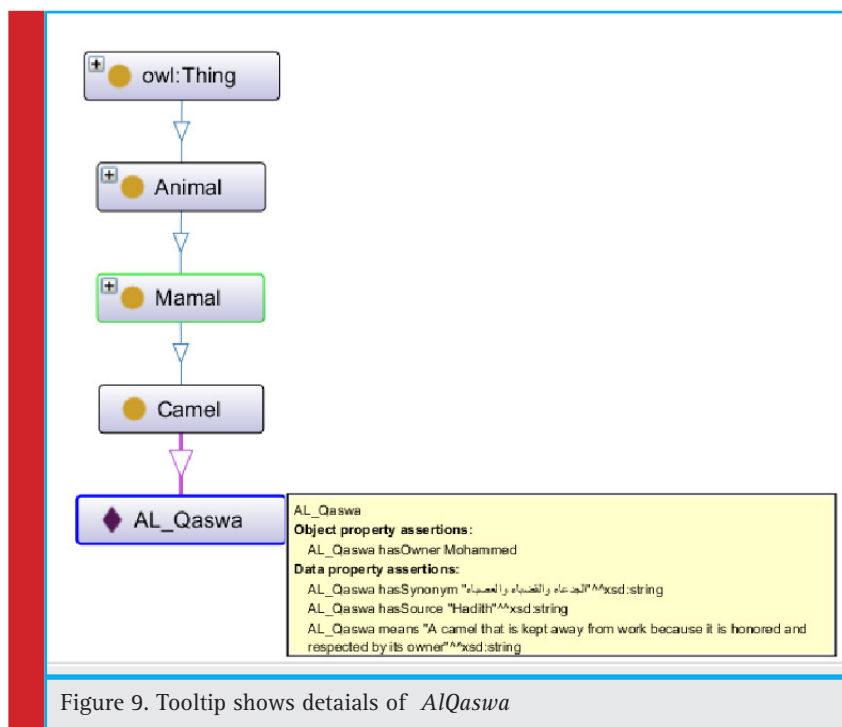


Figure 9. Tooltip shows details of *ALQaswa*

CONCLUSION

This paper aimed to construct an ontology (NEQ-ontology) related to named entities in the holy Quran. The referred named entities are not exhaustive in the holy Quran. For this reason, we have used prophetic traditions (Hadith) to populate the ontology and extract additional named entities. Hadith is considered as a very important source of Islamic knowledge. In our work, we have used the Sahih Al-Bukhari and Sahih Muslim collections.

In future work, we want to investigate methods to (semi)-automatically find not only instances of classes, but also new classes of the ontology. This work can be extended to other Hadiths (al-Tirmidhi, Ibn Maja...) and Fiqh (Temporal interpretation of Sharia rules (Islamic law)). In addition, implementation of complete application covering the entire proposed ontology needs to be accomplished.

ACKNOWLEDGMENT

This work was supported by Taif University. The college of Computer science and Information Technologies.

REFERENCES

- [1] A. B. M. S. Sadi *et al.* Applying ontological modeling on quranic "nature" domain," *2016 7th International Conference on Information and Communication Systems (ICICS)*, Irbid, 2016, pp. 151-155. doi: 10.1109/ IACS.2016.7476102
- [2] K. Dukes & E. Atwell, E. LAMP: A Multimodal Web Platform for Collaborative Linguistic Analysis., in Nicoletta Calzolari; Khalid Choukri; Thierry Declerck; Mehmet Ugur Dogan; Bente Maegaard; Joseph Mariani; Jan Odijk & Stelios Piperidis, ed., 'LREC' , European Language Resources Association (ELRA),2012 , pp. 3268-3275 .
- [3] H. Khan, S. Saqlain, M. Shoaib, and M. Sher, "Ontology Based Semantic Search in Holy Quran," *International Journal of Future Computer and Communication* vol. 2, no. 6, pp. 570-575, 2013.
- [4] M. Bin Ismail Abu Abdullah Bukhari Aljafee. "Sahih Bukhari". Dar tuq alnaja. Edition: I, 1422.
- [5] M. Ibn Al-Hajjaj. "Sahih Muslim". <http://al-hakawati.net/arabic/Civilizations/80.pdf>. Viewed marc. 2017.
- [6] N. F. Noy, and Mcguinness D.L. " Ontology development 101: A guide to creating your first ontology". University of Stanford, Stanford. 2001.
- [7] Corpus quran. <http://corpus.quran.com/ontology.jsp>. Viewed march 2017.
- [8] G. Salton, A. Wong, and C. Yang. "A Vector Space Model for Information Retrieval", *Journal of the American Society for Information Science*, vol. 8, no. 11, pp. 613-620, 1975.
- [9] Protege. <http://protege.stanford.edu/>Viewed march 2017.
- [10] S. Saad, N. Salim and S. Zainuddin, "An early stage of knowledge acquisition based on Quranic text," *2011 International Conference on Semantic Technology and Information Retrieval*, Putrajaya, 2011, pp. 130-136.doi:10.1109/ STAIR.2011.5995777