Computational Ontologies for a Semantic Representation of the Islamic Knowledge

Bendjamaa Fairouz, Taleb Nora

HAL Id: hal-01913891
https://inria.hal.science/hal-01913891

Submitted on 7 Nov 2018

Computational Ontologies for a semantic representation of the Islamic knowledge

Bendjamaa Fairouz and Taleb Nora

Computer science Department, University of Badji Mokhtar Annaba

bendjamaafairouz@gmail.com, talebnr@hotmail.fr

Abstract- In spite of the efforts made in the Arabic language on the syntactic and semantic level, it remains very restricted, even those on the Arabic Sacred Book are few and very limited, due to its difficulties and peculiarities. In this paper we tried to shed the light on some of the recent works that have been conducted to present a semantic representation and manipulation of the Islamic texts to define the problems, limitations and the possible future works that need our intention to improve the semantic support in the Arabic religious texts. Furthermore, we intent to briefly present our project that aims to help us reading, understanding, and interpreting the Islamic legislative sources. The goal of this project is divided into two main tasks which are the creation of an ontology representing the Islamic knowledge and the development of a system which can analyze this knowledge. The ultimate goal is to assist the muftis and facilitate their job.

1 Introduction

Any religion in the word has a number of guides and directives to be followed by its believers, and Islam has no exception. The Islamic laws are derived from the legislative sources which are: Quran (the holy book), Hadith (the words and acts of the Prophet Mohamed peace be upon him), Ijtihad (consensus of the companion) and Quias (the analogical deduction).

Quran is the main religious source of the Islamic rulings or jurisprudence. One of its miracles is its unique style written in classic Arabic. Linguistically and in terms of perfection; it is considered the best Arabic scripture with its expressions and words, in that great meanings were expressed in two or three words and sometimes the same meaning was rehearsed in different and various ways. This holy book contains a huge amount of knowledge in various subjects like: the pillars of Islam, faith, general and political relations, science and

art, holy Quran, organising of financial relations, human and social relations, Al-Jihad, religions, judicial relations, working, stories and history, human and ethical relations, trade, agriculture and industry and call for Allah (Daawah).

Lately, Islamic data knew a huge explosion with the growth of Information technology. One of the most used sources of information to search the Islamic know-ledge is the web. However, this knowledge is sparsely spread all over the web and it is not well organized which make it hard to access, process or reuse it.

Moreover, the majority of software and websites for searching on the Islamic texts are keyword based. And the absence of the semantic support affects the precision of the results, especially when dealing with the ambiguity of the Quranic text and the Arabic language in general.

As most Islamic resources are represented as a plain text or image documents, achieving machine interoperability was a major problem. After the introduction of the semantic web, Ontologies were very popular in different communities such as knowledge management, natural language processing, and information retrieval. Ontologies were defined for the first time in computer science by Gruber an explicit specification of a conceptualization [1]. As an effect, a lot of works have been done to take advantage of the ontologies for a machine-readable representation to semantically model the Islamic domain.

This paper is laid out as follows: Section 2 gives background information on Arabic and Quranic texts difficulties. Section 3 is dedicated to outlining the different main sources of the Islamic religion. Section 4 gives a brief overview of some relevant works on Islamic texts. Section 5 briefly introduces our in-progress project. Section 6 presents our recent results. Section 7 shows some possible future works. Finally, we conclude this paper within the section 8.

2   Arabic Linguistics:

The Arabic script-based languages share in different degrees an explosion of homo-graph and word sense ambiguity. Dealing with such a problem represents a real chal-lenge to NLP systems. Resolving ambiguity in NLP requires representation not only of linguistic and contextual knowledge but also of domain knowledge. Ambiguity in Arabic is enormous at every level: lexical, morphological and syntactic. Another se-rious problem is tokenization and it is extremely common in Arabic to find a token that can function as an entire sentence in English (Fig.1).

In order to facilitate the comprehension of the Islamic sources structure, the Arabic grammar rules are used to define the meaning which is very important since no decisions can be deduced except when the content of these sources are well un-derstood [2].

It is commonly known that Arabic is one of the most difficult languages. In fact, each language has its problems and limitations. In Arabic for instance, it can be the agglutination because as Arabic native speakers we are able to read any text automatically without any agglutination signs, but it can be more

Figure 1: Example of the Arabic language difficulties: the tokenization problem.

Figure 2: Examples of the Arabic language difficulties: the problem of agglutination.

challenging for automatic processing systems or non native-speakers as showsn in the Fig. 2. At the other hand, Arabic language has a very strict grmmar rules which can be helpful in limiting the problems of the automatic processing of Arabic texts. So the problem in this case is the lake of research and works done on the languge rather than the difficulty of the language itself. For exemple, Arabic is spoken by more than 300 million people in over 22 countries, but the works made regarding the automatic processing of Arabic or ontologies are almost non-existent, and a big part of these works are very limited especially compared to the evolutions of other languages. Among these works we can quote the works presented in [3]-[8].

3   Arabic legislative sources

The four legislative sources: the holy Quran, Sunnah of the prophet, consensus of the companion (Ijama) and the analogical deduction (Quias) presented in the Fig.3). Also, we have Fiqh which is the science of having the knowledge of decisions of all Islamic laws which are extracted from the four Islamic legislative sources. In the other hand, we have the foundation of jurisprudence (Usul EL-Fiqh) which it is the theoretical bases relating to the methodology which contains indications and methods used to extract Islamic judgments from the
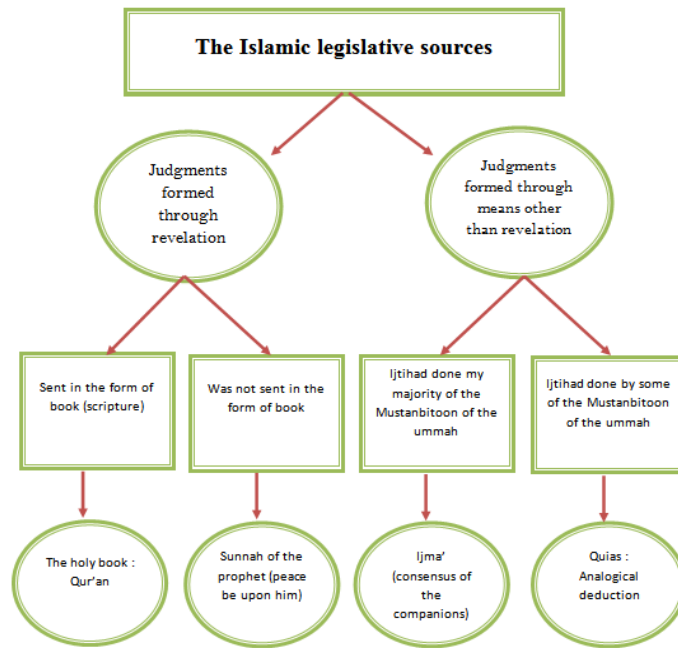
Figure 3: The Islamic legislative sources.

four Islamic legislative sources (Fig. 3).

4 Islamic Ontologies:

4.1 Quranic ontologies:

Many applications based on Quranic text have been built to facilitate information retrieval and knowledge sharing. Some works used the Quran in its original standard Arabic format; others used a translated version like English, Malay etc. Following are some of the recent studies:

Mustapha, A [9] proposed a dialogue-based visualization system called AQI-LAH to facilitate navigating and learning the Quranic text. The source of knowledgebase used was an English version of the Quranic text. This prototype was able to answer users query by listing the related verses.

Fouzi H. et Al [10] based their work on statistic and linguistic methods. They applied a linguistic pattern-based approach to extract the concepts and association rules to extract the conceptual relationships from the Qur'anic text related to the his-torical stories of the prophets.

Al-Yahya, M. et al [11] worked on designing and implementing an ontological model capable of representing the Arabic language lexicons. The application was applied on Time vocabulary in the Holy Qur'an. This ontology only contains

18 concepts where the temporal sequencing of time cannot be captured.

Aliyu R. et al in [12] and [13] proposed a framework capable of responding to complex natural language queries related to historical concepts mentioned in the Quran. They based their work on a knowledgebase containing an annotated ontology in RDF, where users queries are reformulated to match the knowledgebase representation for concept retrieving.

Abbas, N.H. [14] proposed a bilingual search website for the abstract and con-crete concepts in the Quranic text. This website is based on a syntactic search using keywords in the user's query to retrieve the answer according to their occurrence in verses. The limit of such systems is the use of "keyword-based algorithms", and thus they do not provide any semantic or contextual information. To enhance the results she used eight English translations of the Quran and extended the search by using lemmas rather than the exact words of the queries.

Lamraoui Y. [15] proposed a research model incorporating Dukes ontology in various levels of the information search system. He integrated the ontology to represent the Quran and to reformulate the users queries.

Zaidi, S. In [16] worked on the implementation of a new process for building ontologies from Arabic texts, and its application on the Qur'an. She proposed a hybrid method for extracting simple and complex terms, as well as the semantic relations. She used the The Quranic Arabic Corpus proposed by Dukes and the Al-Sulaiti corpus proposed by Eric Atwell

Sharaf, A.B. et al [17] designed a new knowledge representation for the Quranic text. They were able to propose a FrameNet frames for the Arabic verbs mentioned in the Quran. To manage the ambiguity of some cases, several English translations of Quran as well as books of Tafsir were used.

Among the prominent works on Quran, we find the works of K. Dukes, who created a morphological and syntactic annotation for the Qur'an [18]. This project was conducted at Leeds University and their website is world-wide collaboration. Also, he proposed a Quranic ontology in [19] that uses knowledge representation to define the key concepts in the Quran, and shows the relationships between these concepts using predicate logic.

S. Saad and Al [20] -[22], which uses an approach based on a combination of NLP techniques, information extraction and text meaning technologies, to create an ontol-ogy for Islamic concepts in the Quran. But the work presented in [20]; covers only 63 verses related to the obligatory prayers.

A. Hakkoum and Al [23], their ontology covers the following subjects: Quranic chapters and verses, each word of the Quran and its root and lemma to facilitate the keyword search, it does not cover words morphology search but they stated that they will add links to QVOC ontology later on.

4.2   Hadith ontologies

Abdelhamid, M. Et al [24] developed a tool that helps in compiling all the authentic Hadith from the Malay translation of the six books containing the authentic collection of Hadith text (Bukhari, Muslim, Abu Dawud, Tirmidhi,

Nasai, Ibn Madja). The final tool was a well-structured relational database with a user interface.

Yehya M. Et al [25] proposed a decision support system to judge Hadith Isnad using Ontologies. Their work is based on the methodology used by Hadith scholars.

Mohammed G.M. [26] proposed an Ontology-based approach to enhance the process of information retrieval from Al-Shamelah digital library. This work presents a method to support semantic search with complex queries by proposing a new ontology to model concepts from Al-Shamela digital library (ADL). For the evaluation process, they compared the results obtained from their system to the results obtained by the ADL. This system was applied to Hadiths covering the Prophetic medicine domain presented in the ADL.

4.3 Islamic legal rulings Ontologies

The Islamic legal ruling represents the divine law revealed in the Quran and the Sunnah and developed by the consensus of companions (Ijma) and the analogical deduction (Quias). Despite the importance of the Islamic legal rulings, they are not semantically represented. As to our knowledge, no work has been done in this area.

5   Our project

Our project is to develop a new system which aims to extract Islamic judgments with the related evidence texts from Islamic legislative sources. The users can ask any question that requires a deep reasoning using complex natural language. The final application could be used by Muslims, non-Muslims and by the decision-makers in the field of El-Fatwa too.

This work is based on Ontologies representing the Islamic knowledge scattered on the four legislative sources. Nevertheless, the knowledge modeling techniques from an Arabic corpus and the technical analysis of knowledge contained in ontologies are sparsely studied, which requires a deep epistemological research. Even more, the application of such studies on Qur'anic texts is very limited. A new work in this area certainly brings benefits to the Islamic world and the Arabic world in general, as well as a huge support for the progress of the modern science in all its fields. The purpose of this project is to present an interdisciplinary approach, which allows us to correctly read, understand and interpret the Islamic legislative sources.

The work of our project is divided into two major axes:

- The first axe consists of the construction of the ontological model representing the four legislative sources of Islam. Diverse problems are put at this stage, but since many works (some presented in this paper) are done in this domain, they give us the methodological frames for this process. During this phase we create a knowledge base combining different available Quranic ontologies like the Ontology developed at Leeds University.

Whe-reas for the rest of the legislative texts we aim to create our own ontological representation.

- The second axe consists of the analysis phase of the ontology built to supply an answer to a given question. Indeed, the absence of a system which can analyze on ontology and supply a result leads to design a complementary tool for Protg 3.2 to reach the aimed goal.

The final result of this project is a dialogue-based system where users can enquire the system in Arabic. Where needed, the system can ask the user for more details about its query and at the end of the dialogue, the system will generate an answer containing the Islamic judgment with the related verses and evidence texts.

6   Results:

At is stage we aim collecting different existing ontologies to better describe the Islamic knowledge. Until now, we have collected different ontologies and we quote: The Quranic annotation of K. Dukes [19], The Solat Ontology of S. Saad and Al [20], the domain Ontology of A. Hakkoum [23] and our Hadith Ontology.

6.1   Our ontological model:

The automatic processing of texts in Arabic is not very fruitful, due to the complexity of the Arabic language and the lack of tools that allow proper treatment of the language. This has led us to the manual construction of our ontology. By analyzing the texts of the corpus of the Hadith and with the help of the most relevant terms extracted with RapidMiner, we were able to conceive the dictionary of concepts. Then we extracted the different relationships between these concepts, which allowed us to conceive the conceptual model of our ontology. And finally, we have implemented this ontology with the ontology editor Protg (Fig. 4), which we evaluated with some SPARQL queries (fig. 5).

For our ontology we used Volume 1 of Sahih Al-Bukhari which contains the fol-lowing books: book of Revelation, The belief, Ablutions, Menstrual periods, Prayers, The times of prayer, etc. We implemented our ontology with Protg.
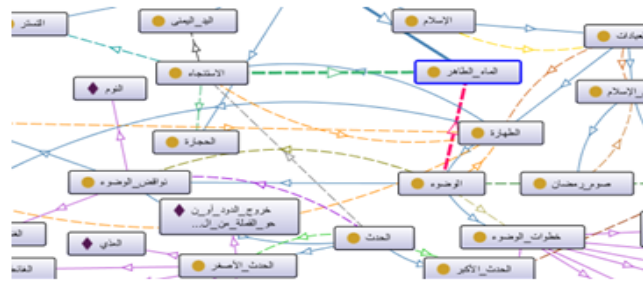

Figure 4: A sample of the hadith ontology graph in Protege.

```
SELECT ?arkan
         WHERE { ?arkan rdfs:subClassOf onto:أركان_الإسلام }
```

الشهادتان
الصلاة
الحج
الزكاة
صوم_رمضان

Figure 5: Result of Sparql query for concepts subsumed by .

7   Future work

- In Quran or Islamic sources in general, we can find different words or phras-es used to refer to the same thing. For this reason, it is preferable to use a large collection of synonyms to help understanding the meaning and avoid-ing the ambiguity.

- Creating Ontology for Tafsir to facilitate the comprehension of the Islamic knowledge.

- Creating a domain ontologies for the Islamic knowledge

- Using Protg with other programming languages and tools to enhance the results.

8   Conclusion

In this paper, we presented some works previously made on the Islamic texts. The extensive efforts made in this field were toward ontology building process. Even though the majority of these works were done on Quranic text only and they remain very limited, they either focus on certain Surah or verses related to a given topic with some simple queries or struggle dealing with the ambiguity of natural language.

   To summarize, there is a huge need to create domain ontologies for the Islamic legislative sources rather than focusing our efforts and time on sample ontologies. Moreover, presenting a semantic representation of the Islamic rulings is a critical need to guaranty the correct understanding the Islamic religion.

References

[1] Gruber T.: A translation approach to portable ontology specifications" Knowledge Acquisition, 5(2):199-220 (1993)

[2] Institute of Islamic banking and insurance, http://www.islamic-banking.com/islamic-jurisprudence.aspx, last visit: 06/05/2016

[3] Nizar H.: Arabic Morphological Representations for Machine Translation, Center for Computational Learning Systems Columbia University.

[4] Al-Agha I.: Using linguistic Analysis to translate Arabic natural language queries to SPARQ , Faculty of Information Technology, The Islamic University of Gaza, Gaza Strip, Palestine.

[5] Jarrar M.: Arabic ontology engineering-challenges and opportunities, Birzeit university Palestine (2011)

[6] Jarrar M., Meersman R.: Ontology Engineering -The DOGMA Approach, Vrije Universiteit Brussel.

[7] Boulaknadel S.: Traitement Automatique des Langues et Recherche d'Information en langue arabe dans un domaine de specialit: Apport des connaissances morphologiques et syntaxiques pour l'indexation, (2010)

[8] Zaidi S., Laskri M., Abdelali A.: Arabic collocations extraction using Gate, in Proceedings of the 2010 International Conference on Machine and Web Intelligence (ICMWI10), (2010)

[9] Mustapha A.: Dialogue-based visualization for Qu'ranic text, European Journal of Scientific Research, (2009)

[10] Fouzi H., Abdullah A.N., Abdullah A.M., Rayan A., Abdulmalik S.: Using associa-tion rules for ontology extraction from a Qu'ran corpus.

[11] Al-Yahya M., Al-Khalifa H., Bahanshal A., Al-Odah I., Al-Helwa N.: An on-tological model for representing semantic lexicons: An application on time nouns in the Holy Qur'an, Arabian Journal for Science and Engineering, (2010).

[12] Aliyu R.Y., Rabiah A.K., Azreen A., Masrah A. M.: Qu'ranic Verse Extraction base on Concepts using OWL-DL Ontology, Research Journal of Applied Sciences, Engineering and Technology, Published: December 15, 2013.

[13] Aliyu R.Y., Rabiah A.K., Azreen A., Masrah A.m.: Semantic Web Application for Historical Concepts Search in Al-Quran, International Journal on Islamic Applications in Computer Science And Technology, Vol. 2, Issue 2, June 2014, 1-7

[14] Abbes N. H.: Qu'ran Search for a concept tool and website, School of Computing, University of Leeds School of Computing, (2009)

[15] Lamraoui Y.: Recherche intelligente des informations dans le coran, Badji Mokhtar university Algeria, (2011)

[16] Zaidi S.: Une plateforme pour la construction dontologie en arabe: Extraction des termes et des relations partir de textes (Application sur le Saint Coran), Badji Mokhtar University Algeria, (2013)

[17] Sharaf A.B., Atwell E.: A corpus-based computational model for knowledge representation of the Quran, University of Leeds, (2009)

[18] Dukes K., Atwell E., Sharaf A.B: Syntactic annotation guidelines for the Quranic Arabic Dependency Treebank, in Proceedings of the Language Resources and Evaluation Conference (LREC), (2010)

[19] Dukes K.: Ontology of Quran Concepts, available at http://corpus.Quran.com/ontology.jsp, (2012), last visit: 04/05/2016,

[20] Saad S., Salim N., Zainal H., Muda Z.: A process for building domain ontology: An experience in developing Solat ontology, in Proceedings of the Inter-national Conference on Electrical Engineering and Informatics, Bandung, Indonesia, (2011)

[21] Saad S., Salim N.: Methodology of Ontology Extraction for Islamic Knowledge Text

[22] Saad S., Salim N., Zainal H.: Islamic knowledge ontology creation", in Pro-ceedings of the Internet Technology and Secured Transactions, (2009)

[23] Hakkoum A., Raghay S.: Ontological approach for semantic modeling and query-ing the qur'an. In Proceedings of the International Conference on Islamic Applications in Computer Science And Technology, 2015.

[24] Abdelhamid Y., Mahmoud M., El-Sakka T.M.: Towards Enhancing the Compilation of Al-Hathdith Text in Malay, Quran and Hadith Research Group, University of Malaya.

[25] Yehya M. D.: An Ontology-Based Approach to Support the Process of Judging Hadith Isnad, Master thesis, IslamicUniversity of Gaza, (2013)

[26] Ghazi A.M: An Ontology Based Approach to Enhance In-formation Retrieval from Al-Shamelah Digital Library, Master thesis, IslamicUni-versity of Gaza, (2015)