

An Ontological Model of Hadith Texts

Semantic Representation of Hadith

Bendjamaa Fairouz¹, Taleb Nora², Arari Amina Nouha³

Complex System Engineering Laboratory
Department of Computer Science
Badji Mokhtar University
Annaba, Algeria

Abstract—The Hadith being the second source of legislation after the Holy Qur'an in the religion of Islam, it represents a large body of knowledge in unstructured textual form. The specification of Hadiths makes its automatic exploitation a rather robust and an almost impossible task. To enable different types of computer systems to exploit this knowledge, various researchers used a formal representation of the semantics of Hadith. The widely used semantic representation is ontology defined as concepts and relations extracted from the Hadith in the form of a structure interpretable both by the machine and the human. In this article, we propose an ontology of the Hadith using an approach inspired by the "METHONTOLOGY" methodology. In this project, we are dealing with religious texts in traditional Arabic, and we face many difficulties in achieving complete precision and correctness. Hence, we decided to follow an entirely manual process to ensure the correctness of the results. Since manual ontology development is both time and effort consuming, we decided to focus only on "Wudhu2" related Hadiths.

Keywords—*Ontology engineering; Islamic ontology; Methontology; semantic representation*

I. INTRODUCTION

To be able to exploit a given domain knowledge, we must explicitly represent it to make it directly used by applications.

Ontology engineering was born of the need for knowledge representation. Ontologies' aim is representing knowledge in a way that can be interpreted by both man and machine. An ontology is a set of concepts and relations that constitute the knowledge of a domain.

Building ontologies is the process of transforming the most relevant knowledge in a domain into a structure that will allow the automatic exploitation of this knowledge.

The Hadith is the set of words and deeds of the prophet Mohamed (PBUH) expressed by terms used in the spoken language of Arabic. For Muslims, Hadiths are the second source of Islamic legislation after the Koran. The collections of the Hadiths are very voluminous. Therefore, the extraction of information is time-consuming, especially when taking into consideration a large number of Hadiths consulted for each query. Hence, the urgent need to benefit from the knowledge representation formalisms.

The automatic processing of the Arabic language is considered difficult to apprehend because of morphological

and structural characteristics of this language, such as polysemy, irregular forms of certain words, and derivative properties.

The objective of this project is the construction of an ontology-based on Arabic texts and, more specifically, an ontology that represents the semantics of the Hadith text in its original form.

This article is structured as follows: the following section dedicated to basic notions about ontologies and similar work. The third section details the proposed approach to build the Hadith ontology. The fourth section presents the results obtained. Finally, we conclude this paper by the prospects for future work.

II. STATE OF THE ART

A. What is an Ontology?

In the literature of AI, we find many definitions of ontology, the most commonly used and referenced proposed by Gruber in [1] that was later refined by Borst as "An ontology is a formal and explicit specification of a shared conceptualization." [2].

B. Constituents of an Ontology

Ontologies are representations of knowledge, containing terms and statements that specify the semantics of a given domain of knowledge within a given operational framework. [3].

The main constituents of a given ontology are:

- Concepts, also called terms, represent a principle, an idea, or an abstract notion that is semantically evaluable and communicable. [4].
- Relations, which represent a type of interaction, or associations that exist between the concepts of a domain. [4].
- Instances are individual nodes in a semantic network, representing individual objects of the field of interest, such as a car or a specific person. [4].
- Axioms used to model sentences that are always true [5]. Often expressed in the logic of first-order predicates.

C. Methontology

There are many methodologies for constructing ontologies in literature. The process of ontology development refers to the different activities that one accomplishes to obtain an ontology. For this project, we chose METHONTOLOGY, proposed by M. Fernandez et al. in [6], which allows us to construct ontologies using an intermediate conceptual model, without requiring any prior knowledge of domain concepts.

The main phases of Methontology are:

- Specification of the purpose of the ontology, its end-users, its scope and the set of terms to be represented, the sources of knowledge [6].
- Conceptualization: in which we structure the knowledge of the domain in a conceptual model using the vocabulary already defined in the specification phase. [6].
- Implementation: The result of this phase is an ontology coded in a formal language such as CLASSIC, Ontolingua, Prolog [6].
- Evaluation: is to form a technical judgment on the ontology using the specification document realized in the first step. [6].

D. Similar Work

Many works of construction and exploitation of ontologies have been carried out in different fields. In this paper, we are interested in ontologies in the domain of Islam, and more specifically, the Hadiths.

A given Hadith has two main parts: the narrative or the content part of the Hadith is called Matn, and the chain of narrators (reporters) through which the narration was transmitted and then recorded, is known as Sanad or the chain of narrators. The Sanad plays the most crucial role in determining the authenticity of the Hadith, which is the most crucial indicator of whether to accept or reject a Hadith.

Azmi A.M. and Bin Badia N. in [7] is an ontology named "HadithRDF" that is used to represent the chain of narrators in a standard format and then graphically represent its complete tree. HadithRDF is designed to cover a large number of Hadith books such as Sahih El-Bukhari in the Hadith corpus.

Basharat et al. in [8] present the structure of the Hadith, and then, based on this structure, they propose a conceptual model of ontology for the Hadith. In this model, the Matn and Sanad are represented as separate entities related to the entity Hadith by the relation "part of" and "hasMatn." Also, they represented the level of Hadith's authenticity, chapters, books, and the collection to which it belongs.

Al-Rumkhanian A. et al. in [9] proposed a Hadith ontology for Prophetic medicine 'الطب النبوي'. They used authentic Hadiths as a corpus, and they proposed as future work to further extend the ontology to generate treatments for some diseases according to Prophetic medicine automatically.

Harrag F. et al. in [10] proposed a Hadith ontology based on Sahih El-Boukhari. They used association rules to extract the relations between the concepts in Sahih El-Boukhari.

Al-Masri M.G. in [11] proposed a new ontology to model concepts from Al-Shamela digital library (ADL). The ontology covers the Prophetic Medicine domain. For the evaluation, they compared their results to the results obtained by the ADL.

Hadith was inspiring some researchers to apply different techniques for knowledge modelling or information retrieval to process it, we quote:

Harrag F. et al. in [12] used text mining techniques to extract Islamic knowledge from Hadith. They used the vector space model, term frequency, cosine measure, and inverse document frequency. This tool retrieves Hadiths classified by similarity degree to the user's query.

Moath N. et al. in [13] used the head-driven phrase structure grammar formalism (HPSG) to describe the lexicon for Hadith. The final lexicon is a set of XML documents. This is a corpus-based project; they used a corpus from Al-Bukhari and Muslim books.

III. THE APPROACH OF CONSTRUCTING THE HADITH ONTOLOGY

The proposed approach to ontology construction consists of the steps outlined in Fig. 1.

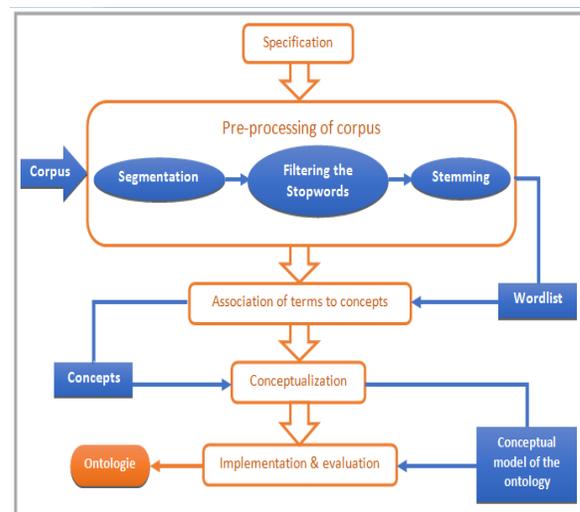


Fig. 1. The Proposed Process for Constructing the Hadith Ontology.

A. Specification

The specification step is the same as in the "Methontology" methodology, where we must specify the purpose of the ontology, the users of this ontology, its scope, the source of knowledge (corpus), and the domain.

- Domain: Islamic knowledge and more specifically, the Hadith which is the set of words and acts derived from the prophet of Islam Mohamed.
- Objective: The semantic representation of the Hadith to be used for the automatic extraction of Islamic laws or laws.
- End-users: Any developer who plans to build a system based on the automatic extraction of Hadith text.

- Source of knowledge (corpus): Volume 1 of the collection of Hadith of El-Bukhari "Sahih El-Bukhari."
- Scope: This aspect consists of determining a priori the list of the most important terms that will contain the ontology, among these terms: العبادات, أركان الإسلام, and more.

B. Pre-Processing of the Corpus

After specifying the source of knowledge, which is, in this case, in unstructured textual form, we eliminate unnecessary information such as titles and subtitles, as well as the chain of narrators of each Hadith, keeping only the text or Matn of the Hadiths. Then we perform the following tasks using RapidMiner:

1) *Segmentation or tokenization*: To divide the input text files and to extract their contents in the form of separate words.

2) *Filtering the Stop-words*: To keep only the relevant words to the studied domain.

3) *Stemming*: After filtering the stop-words, we use the operator Stemming (Arabic) to transform the words into their root or radical. This step will minimize the number of extracted terms.

At the end of this step, we obtain a list of the relevant terms (wordlist) with their frequency of occurrences in Table I.

C. Association of Terms to Concepts

Using the wordlist and their occurrences obtained in the previous step, we construct the list of concepts with the help of an expert in the field of Islamic law. This step was performed manually.

The extracted list contains the concept, its description and its synonyms (Table II).

D. Conceptualization

After having collected the set of concepts that will be included in the ontology specified their definition in natural language and identified their synonyms. We organize and structure the domain knowledge using a set of intermediate, semi-formal representations in the form of tables and graphs. We perform the following tasks.

1) *The construction of taxonomy*: The task of defining the taxonomic relations between the concepts, we classify the concepts collected previously hierarchically (Concept child / Concept Parent). In Table III, an example of some retrieved taxonomic relations.

2) *The identification of binary relations that link two concepts together*: We construct the table of relations which contains the name of the relation, the source concept and the target concept as presented in Table IV.

3) *Describing in detail each attribute*: Attributes are properties that take their values in the predefined types (String, Integer, Boolean, Date and others).

TABLE I. A SAMPLE LIST OF TERMS EXTRACTED WITH RAPIDMINER

Term frequency	Total occurrence	Word
0.008446717	9	نوم
0.008446717	9	حدث
0.008446717	10	ولي
0.008446717	10	مضمض
0.008446717	10	ماء

TABLE II. A SAMPLE LIST OF CONCEPTS

Synonyms	Description	Concept
/	الخضوع لله، الدين الذي بعث الله به محمداً صلَّى الله عليه و سلم	الإسلام
الشعائر الدينيَّة	طاعة الله المُتمتَّلة بامتثال ما أمر به و أخبر به عنه الرِّسَل و الأنبياء	العبادات
/	العبادة المخصوصة المبيَّنة حدود أوقاتها في الشَّرِيعَة	الصلاة
الوضوء الأكبر	هو غسل و تطهير البدن بإفاضة الماء عليه لاستباحة الدخول في العبادات	الغسل
الوضوء الأصغر	غسل أعضاء معينة و المسح عليها بالماء الطاهر مع النية	الوضوء
/	النَّجاسة الخُكمية التي ترتفع بالوضوء أو الغُسل أو التَّيْمُم	الحدث
...

TABLE III. TAXONOMIC RELATIONS

Parent	Concept
Thing	الإسلام
الإسلام	العبادات
العبادات	الصلاة
العبادات	الطهارة
الطهارة	الغسل
الطهارة	الوضوء
...	...

TABLE IV. A SAMPLE OF THE RELATIONS TABLE

Target concept	Source concept	Relation
العبادات	الإسلام	يُتضمن
أركان الإسلام	العبادات	تشمل
العمرة	العبادات	
الطهارة	العبادات	جزء من
أركان الإسلام	الصلاة	
الطهارة	الغسل	عبارة عن
الطهارة	الوضوء	
...

Table V presents sample attributes which contain: the name of the concept, the type and the cardinality of each of its attribute.

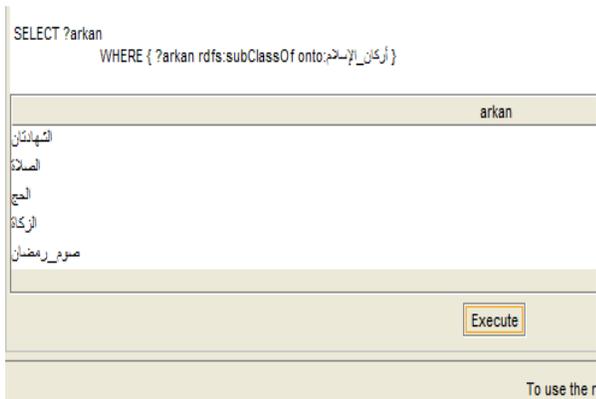


Fig. 4. The Result of a SPARQL Query for Concepts Subsumed by أركان الإسلام.

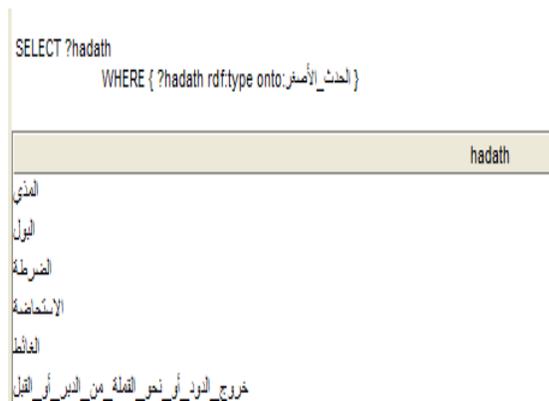


Fig. 5. The Result of a SPARQL Query for Instances of الحدث الأصغر.

V. FUTURE WORKS

To our knowledge, no similar works have been done until this day. The previous works either use the Hadith text in a different language than Arabic or present a domain knowledge representation.

At the end of this research, the results can be used by Muslims, non-Muslims, scholars, and Hadith experts.

The final project can be combined with other ontologies modelling the different Islamic books to present a full semantic representation of Islamic knowledge. Also, it can be used for semantic search and information extraction tools.

VI. CONCLUSION

This project consists of the construction of an ontology that represents the semantics of the Hadiths and the knowledge that can be extracted from these voluminous textual sources of knowledge in Arabic.

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. These limitations have led us to the manual construction of the ontology. By analyzing the Hadiths and 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 the ontology. Finally, we have implemented this ontology with the ontology editor Protégé, which we evaluated with some SPARQL queries.

REFERENCES

- [1] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993, doi: 10.1006/knac.1993.
- [2] W.N. Borst, "Construction of engineering ontologies for knowledge sharing and reuse," University of Twente, Enschede, Centre for Telematica and Information Technology, 1997.
- [3] F. Furst, "L'ingénierie ontologique," rapport technique, Institut de recherche en Informatique de Nantes, 2002.
- [4] S. Grimm, A. Abecker, J. Volker, R. Studer, "Ontologies and the Semantic Web," In *Handbook of Semantic Web Technologies*, Springer Berlin Heidelberg, pp. 507-579, 2011.
- [5] A. Gomez Pérez, V.R. Benjamins, "Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods," *Proceeding of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5)*, Stockholm, Suède, pp. 1.1-1.15, 1999.
- [6] M. Fernandez, A. Gomez-Pérez, N. Juristo, "METHONTOLOGY: From Ontological Art Towards Ontological Engineering", *AAAI Technical Report SS-97-06*, Polytechnic University of Madrid, Spain, 1997.
- [7] A.M. Azmi, N. bin Badia, "e-narrator – an application for creating an ontology of hadiths narration tree semantically and graphically", *The Arabian Journal for Science and Engineering*, Vol. 35, No. 2C, 2010.
- [8] A. Basharat, B. Abro, I.B. Arpinar, K. Rasheed, "Semantic Hadith: Leveraging Linked Data Opportunities for Islamic Knowledge", *Conference: Linked Data on the Web, LDOW2016*, Montreal, Canada, 2016.
- [9] Al-Rumkhanian, A., Al-Razganb, M., Al-Farisb, A.: *TibbOnto: Knowledge Representation of Prophet Medicine (Tibb Al-Nabawi)*. In: *Symposium on Data Mining Applications, SDMA2016*, Saudi Arabia, 2016.
- [10] Harrag, F., Alothaim, A., Abanmy, A., Alomaigan, F., Alsalehi, S.: *Ontology Extraction Approach for Prophetic Narration (Hadith) using Association Rules*. In: *International Journal on Islamic Applications in Computer Science And Technology*. 2013.
- [11] Al-Masri, M.G.: *An Ontology-Based Approach to Enhance Information Retrieval from Al-Shamelah Digital Library*. In: *Master thesis, Islamic University-Gaza*. 2015.
- [12] Harrag, F., Cherif, A., Qawasmeh, E.: *Vector Space Model for Arabic Information Retrieval Application to Hadith Indexing*, IEEE. 2008.
- [13] Moath, N., Abdelkarim, A., Musab, A., Abdelrahman, O.: *A Lexicon for Hadith Science Based on a Corpus*. In: *(IICSIT) International Journal of Computer Science and Information Technologies*, Vol. 6 (2), pp. 1336-1340. 2015.