Towards a Joint Ontology of Quran and Hadith

Shatha Altammami[1,2,a], Eric Atwell[2,b] Ammar Alsalka[2,c]

[1]King Saud University (KSU), Saudi Arabia

[2]University of Leeds, UK

[a]Scshal, [b]E.S.Atwell, [c]M.A.Alsalka @leeds.ac.uk

Abstract

The Quran and the Hadith are two religious texts used by Muslims as reference to perform their daily religious obligations. Hence, developing a computational tool to better understand and link these different texts would be useful for Islamic scholars, learners, and laymen. We hypothesise that linking them using a knowledge-based approach by extending a Quran ontology to cover the Hadith is possible. However, there are several Quran ontologies and none of these were rigorously evaluated using a standard ontology evaluation method. For this reason, we started our experiment by enumerating and discussing the existing Quran ontologies. Then the best candidates of the enumerated ontologies are evaluated using a corpus-based approach to visualize the overlap between these ontologies and the Hadith. Our experiment shows that one Quran ontology could be used as a starting point for a larger scope of an Islamic ontology that covers both the Quran and Hadith.

Keywords: Quran, Hadith, Ontology, Ontology Evaluation, Arabic NLP.

1. Introduction

The fascinating idea of capturing human expert knowledge in a machine-readable format is a common goal of ontologies, which is defined by Gruber as 'a specification of a conceptualization'(Gruber, 1991). The word 'conceptualization' refers to the abstract and simplified description of an area of concern one wishes to represent for a specific purpose. Hence, the ontology artifact identifies concepts and relationships to infer associations that form the basis of complex semantic knowledge. In other words, a domain ontology is the backbone knowledge of a particular field, or an expert's knowledge captured in a machine-readable format. This captivating idea sparked research interests in knowledge representations of Islamic teachings.

In the past decade, there have been several attempts to build ontologies that grasp the structure and semantic relationships within and between the verses of the Quran (the holy book of Islam)(Alqahtani & Atwell, 2018). Researchers focused on several knowledge representation structures, with some emphasis on data aggregation to merge the various Quran ontologies. However, no study, to our knowledge, has investigated using these ontologies as the basis of creating an ontology for Hadith, which is the set of narratives reporting the words, teachings, and actions of the prophet Muhammad. We speculate that although the Quran and the Hadith are different, one ontology could be used for both since they cover the same domain of Islamic teachings. Thus, in this study, we investigate which Quran ontology is most suitable for Hadith.

In the following paragraphs, we give a brief background on the Quran, Hadith, and the difference between them. Then enumerate the existing Quran ontologies and discuss the most appropriate candidates for our investigation. After that, we review the ontology evaluation frameworks and discuss the approach we will undertake. Then the experiment of finding the 'fit' of the Quran ontology for Hadith is analyzed. Finally, we explore the possible ways of extending the Quran ontology to cover Hadith and form the Islamic ontology.

2. Quran and Hadith

Muslims believe the Quran is God's divine words transmitted to the prophet Muhammad by the angel Gabriel over a period of 23 years. This holy book enjoins Muslims to follow the guidance of Prophet Muhammad in their laws, legislation, and moral guidance. The clear instruction to emulate the prophet and follow his judgments is necessary because not all Islamic laws and regulations are mentioned in the Quran. Most laws and legislations are obtained from Hadith due to its larger scope and incorporated details. Nevertheless, the Quran is the criterion and standard reference for judging if a Hadith is accepted as authentic by ensuring its agreement with the Quran and not contradicting it. Consequently, finding the links between the two texts is necessary.

Although the Quran and Hadith cover the same domain of Islamic teachings and both are considered classical Arabic, they are distinctive in structure and style. The Quran has 114 chapters, each includes several verses where some consecutive ones are semantically related. On the other hand, each Hadith is a stand-alone statement or act by the prophet that was later written, collected, and compiled into books where scholars have categorized them into topics by relying on their deep knowledge and understanding of Hadith.

The Hadith structure is especially different in that it consists of two parts, as shown in Figure 1. The Isnad is shown in bold, representing the reverse chronological chain of scholars who verbally transmitted the Hadith before it was written into books. The Isnad is usually followed by Matn, which is the actual teaching or prophetic words.

حَدَّثَنَا يَحْيَى بْنُ بُكَيْرٍ، حَدَّثَنَا اللَّيْثُ، عَنْ عُقَيْلٍ، عَنِ ابْنِ شِهَابٍ، قَالَ أَخْبَرَنِي أَنَسُ بْنُ مَالِكٍ، أَنَّ رَسُولَ اللَّهِ صلى الله عليه وسلم قَالَ " مَنْ أَحَبَّ
أَنْ يُبْسَطَ لَهُ فِي رِزْقِهِ، وَيُنْسَأَ لَهُ فِي أَثَرِهِ، فَلْيَصِلْ رَحِمَهُ ".

Yahya bin Bakir told us that Alith told him from Aqeel from Ibn Shihab who said Anas bin Malik told me that the prophet peace be upon him (PBUH) said, "Whoever loves that he be granted more wealth and that his lease of life be prolonged then he should keep good relations with his kith and kin."

Figure 1: Hadith example, Isnad in bold followed by Matn.

If the main structure of the two books is overlooked where the Quran verses are compared to the Hadith Matns, differences still exist in their stylometry. This was concluded by an authorship attribution study using text-mining-based investigation, which shows interesting observations (Sayoud, 2012). For example, unlike the Hadith, the Quran almost always has termination similarity between the neighbouring verses, such as the same final syllable or letter. Another interesting difference is that Hadith uses shorter words than the Quran. Their most significant finding is that 62% of the Bukhari (The most famous Hadith book) Hadith words do not occur in the Quran, and 83% of the Quran words do not occur in the Bukhari Hadith. Despite the significantly different words used in both texts, we hypothesize that finding semantic similarity within and between the Quran and Hadith is possible using ontologies.

Therefore, in this study, we investigated whether it is possible to build a joint ontology of the Hadith and Quran using one of the existing Quran ontologies.

3. Existing Quran Ontologies

A number of Quran ontologies have been developed; some cover the whole Quran while others focus on specific topics. Although this research area was explored for a decade, it is not yet mature with a unified, validated, available Quranic ontology that covers the whole Quran with its various topics. In the following paragraphs, we enumerate the existing Quran ontologies and discuss the most appropriate candidate to be the base of an Islamic ontology that covers the Hadith.

The most cited Quran ontologies aim to cover the whole Quran and were developed at Leeds University (Alqahtani & Atwell, 2018). These were at different levels and for different purposes; including verse level to encode the concept meaning of each verse which was done by Abbas (Abbas, 2009), incorporating more than 1,000 concepts linked to all verses in the Quran. She developed this using an Islamic exegesis textbook "Mushaf Al-Tajweed", where the meaning (or topic) of each Quran verse is elucidated by scholars and encoded in the index.

The other widely used ontology is the Quranic Arabic Corpus (QAC) developed by Dukes where he used the exegesis textbook "Tafsir Ibn Kathir" to elucidate 300 concepts (Dukes, 2015). Similarly, "Tafsir Ibn Kathir" exegesis textbook has been used by Sharaf to build Qursim ontology that consists of 7600 related verses (Sharaf & Atwell, 2012b). Sharaf also introduced QurAna ontology that describes pronoun antecedents in the Quran, which consists of 1050 concepts and 2700 relations that were manually built during his Ph.D. (Sharaf & Atwell, 2012a).

Other researchers created ontologies for specific domains of the Quran. For example, Khan et al presented an ontology based on an Islamic book describing animals mentioned in the Quran (Ullah Khan, Muhammad Saqlain, Shoaib, & Sher, 2013). Another ontology that focuses on a specific domain, presents a model that captures the meaning of Time nouns in the Quran (Al-Yahya, Al-Khalifa, Bahanshal, Al-Odah, & Al-Helwah, 2010). A similar approach was taken by Alromima et al to build an ontology dedicated to Place nouns in the Quran (Alromima, Moawad, Elgohary, & Aref, 2015). Their work is based on "Alam Almakan, Fe Al Quran" which is an MSc thesis of a linguistic computational study. An interesting model to build Quran ontology specific to human relations concepts was introduced by Tashtoush et al (Tashtoush, Al-Soud, Abujazoh, & Al-Frehat, 2017). For each class, they defined a set of synonyms in Arabic, then manually linked the concepts or classes to each verse. The Semantic Quran (Sherif & Ngonga Ngomo, 2009) focuses on concepts in terms of Quran text structure parsing: every chapter, verse, word and lexical item is treated as a "concept", linked by ontology relations such as "isPartOf".

Although the above ontologies are built using different technologies, there were some attempts to unify them (Alrehaili & Atwell, 2018). However, not all ontologies are designed to be reused. Many proposed ontologies are not presented in an applicable and reusable way, but rather to prove a concept in the research environment (Omar et al., 2013).

Some researchers went beyond aligning ontologies by not only combining ontologies, but extending them. Hakkoum et al (Hakkoum & Raghay, 2016) integrated Semantic Quran (Sherif & Ngonga Ngomo, 2009) and QAC ontologies using Owl property *Same-as* to indicate the two

concepts represent the same thing and can be used interchangeably. Then, extend it by covering one topic at a time by manually extracted concepts from Tafsir exegesis books.

The discussed ontologies are further summarized in Table 1, which illustrates the four ontologies attempts to cover all topics across the whole Quran. However, the ontology introduced by Hakkoum (Hakkoum & Raghay, 2016) is composed of the ontology by Dukes (Dukes, 2015) and the one by Sherif (Sherif & Ngonga Ngomo, 2009). Therefore, we will only evaluate the two distinctive ontologies, Abbas's Qurany ontology (Abbas, 2009) and the Quran ontology by Hakkoum. To avoid any confusion between the names of the evaluated ontologies, we will refer to the one by Abbas as ontology 'A' and the one by Hakkoum as ontology 'B'.

Table 1: Existing Quran ontologies.

Paper	Ontology Name	Source	Language	Topics	Format	Number of Concepts	Available
(Dukes, 2015)	QAC	Ibn Kathir	En,Ar	All Quran	XML	300	Yes
(Hakkoum & Raghay, 2016)	Quran	QAC, Semantic Quran	En, Ar	All Quran	OWL	1181	Yes
(Abbas, 2009),	Qurany	Mushaf Al-Tajweed	En, Ar	All Quran	XML	1100	Yes
(Sharaf & Atwell, 2012b)	QurSim	Ibn Kathir	Ar	Similar Verses	SQL		Yes
(Sharaf & Atwell, 2012a)	Qurana	Ibn Kathir	Ar	Pronoun Anacedent	SQL	1,050	Yes
(Ullah Khan et al., 2013)	N/A	Hewanat-E-Qurani,	En	Animals	OWL	167	No
(Al-Yahya et al., 2010)	N/A	Arabic dictionaries and lexicons	Ar	Time Nouns	OWL	53	Yes
(Alromima et al., 2015)	N/A	Alam Almakan, Fe Al Quran	En, Ar	Places Nouns	OWL	99	No
(Tashtoush et al., 2017)	N/A	Tafsir al-Jalalayn	Ar,	Social Relations	OWL	415	No
(Saad, Salim, Zainal, & Muda, 2011)	Solat	Quran + Hadith	En	Solat	OWL	48	No
(Sherif & Ngonga Ngomo, 2009)	Semantic Quran	Tanzil project, QAC	42 languages	Text structure parsing	RDF	26,735	Yes

4. Ontology Evaluation Methods

To empirically measure which Quran ontology fits well as the basis for an Islamic ontology that covers the Hadith, we explore the field of ontology evaluation. There is a plethora of research in ontology evaluation methods and techniques (McDaniel & Storey, 2019). The consensus is that ontology evaluation techniques can be grouped into four categories: gold-standard-based, corpus-based, task-based and criteria-based approaches. Each approach evaluates the ontology differently as shown in the Table 2, where a darker colour in the table represents a better coverage for the corresponding criterion.

The most appropriate approach to evaluate the Quran ontologies 'fit' for Hadith is the corpus-based, for two reasons. First, they are created based on a textual corpus, the Holy Quran book. Second, we are evaluating these ontologies for the Hadith, which is also a textual corpus. The other approaches could be more useful for other types of ontologies that were built using knowledge elicitation from experts.

Table 2: An overview of ontology evaluation methods (Raad & Cruz, 2015).

	Gold	Corpus	Task	Criteria
Accuracy				
Completeness				
Conciseness				
Adaptability				
Clarity				
Computational Efficiency				
Consistency				

A corpus-based approach, also known as data driven approach, is used to evaluate how far an ontology covers a given domain (Raad & Cruz, 2015). A widely used method of corpus-based evaluation is established by Brewster et al (Brewster, Alani, Dasmahapatra, & Wilks, 2004). It relies on three main steps. First, extracting keywords from a corpus either using clustering methods or topic modelling. Then, applying query expansion to capture synonyms using Wordnet or other information retrieval (IR) techniques. Finally, the ontology is evaluated by mapping the set of keywords identified in the corpus to the concepts of an ontology being evaluated.

5. Experiment

In this experiment, we evaluate two ontologies to investigate their appropriateness for being the basis of an Islamic ontology that covers the Hadith in addition to the Quran. Table 3 below shows a Quran verse with the associated concepts extracted from the two ontologies. It is worth noting that Ontology B not only captures topics, but also the structure of the Quran and the links between similar verses. Moreover, it includes the Tafseer (exegesis) extracted from two books, Jalalayn and Muyasser.

Table 3: A Quran verse and concepts extracted from the two ontologies.

	Arabic	English Translation
Quran verse (2:27)	الَّذِينَ يَنقُضُونَ عَهْدَ اللَّهِ مِن بَعْدِ مِيثَاقِهِ وَيَقْطَعُونَ مَا أَمَرَ اللَّهُ بِهِ أَن يُوصَلَ وَيُفْسِدُونَ فِي الأَرْضِ أُولَـئِكَ هُمُ الْخَاسِرُونَ	Those who break Allah's Covenant after ratifying it, and sever what Allah has ordered to be joined (as regards Allah's Religion of Islamic Monotheism, and to practise its legal laws on the earth and also as regards keeping good relations with kith and kin), and do mischief on earth, it is they who are the losers.
Ontology A (Abbas, 2009)	أركان الإسلام، التوحيد	Pillars of Islam, Islamic
	وعيد المفسدين	The Threatening of the Mischievous
	والمجرمين والفاسقين	The Criminals and the Impious
	العمل، العمل الطالح	Action (Work), Bad Deeds
	وعيد المفسدين	Threatening the Spoilers
	الأخلاق الحميدة	Good Morals
	الأخلاق الذميمة	The Dispraised Morals
	الوفاء بالعهد	Keeping Promise
	الإفساد، الخبث، نقض العهد	Corruption, Cunning، Break Promise
	الإنسان والعلاقات الاجتماعية	Human and The Social Relations
	صلة ذوي القربى	Kinship

5

Ontology B (Hakkoum & Raghay, 2016)	أركان الإسلام، التوحيد	Pillars of Islam, Islamic
	وعيد المفسدين	The Threatening of the Mischievous
	والمجرمين والفاسقين	The Criminals and the Impious
	العمل، العمل الطالح	Action (Work), Bad Deeds
	وعيد المفسدين	Threatening the Spoilers
	الأخلاق الحميدة	Good Morals
	الأخلاق الذميمة	The Dispraised Morals
	الوفاء بالعهد	Keeping Promise
	الإفساد، الخبث، الفساد، نقض العهد	Corruption, Cunning ،Break Promise
	الإنسان والعلاقات الاجتماعية	Man and The Social Relations
	صلة ذوي القربى	Kinship
	ايه27، سورة2، حزب1، صفحة5	Verse27, Chapter2, Hizb1, Page5
	متشابه بشدة	Strongly similar (link to one verse)
	متشابه قليلاً	Slightly similar (links to five different verses)
	تفسير الميسر، تفسير الجلالين	DescByJalalayn, DescByMuyaser

Since we are undergoing a corpus-based evaluation of the ontologies, we use the Bukhari Hadith book as our evaluating corpus. This book is the most famous Hadith compilation and it organizes and categorizes Hadiths under headings that indicate the legal implication or ruling the scholar (Bukhari) derived from the subsequent Hadiths based on his deep knowledge and expertise of the Hadith.

The first step in the corpus-based approach for ontology evaluation is keyword extraction. However, instead of extracting keywords from the Hadith Matn (the part of the Hadith that contain the actual statements), we opt to use the Bukhari-Hadith-book section-headings. We speculate that since the evaluated Quran ontologies are based on books like "Mushaf Al-Tajweed" where scholars categories Quran verses into topics, section-headings in the Bukhari's book are sensible keywords of the topic of Hadiths. It is worth noting that some of the Hadith contain verses from the Quran. Hence some scholars might suggest these verses should be excluded from the Hadith text. However, for corpus linguistics research, it is standard practice to keep the source text without editing it. This provides source data for research in overlap or reuse of text.

We hypothesis that there is a larger degree of overlap between the ontology concepts and Bukhari book section-headings than the actual Hadiths Matn and Quran verses. This assumption is based on the study mentioned earlier in Section 2 (Sayoud, 2012). Moreover, prophet Muhammed style in Hadiths statements (Matn) is frequently filled with hyperbole (Brown, 2017), which are dramatic statements different from their evident meaning. For example, the Hadith Matn in Table 4 does not literally mean that such person is not considered a Muslim, but it means a certain action or characteristic is not the conduct of a good Muslim which is captured in the associated section-heading.

Table 4: Example of Hadith Matn and the associated section-heading.

	Arabic	English Translation
Hadith Matn	"منْ غشَّنا فليسَ منَّا"	"One who cheats is not from among us."
Section-Heading	النهي عن الغش والخداع	Prohibition of deceiving others

5.1 Data Sources

In this section, we specify the sources of the data used in the experiment. Ontology A[1] and ontology B[2] can be accessed from their dedicated websites. While the Hadith section-headings were obtained from the LK-Hadith corpus (Altammami, Atwell, & Alsalka, 2019), which is a parallel corpus of the Arabic and the English translation of Hadith. This corpus consists of the six canonical Hadith books where every Hadith component and meta data is annotated. Using it, the researcher can extract specific information; for example, the Hadith Isnads, Matns, or section-headings. This corpus was built automatically using the Hadith segmentation tool that was created to deconstruct every Hadith to its two main components, Isnad and Matn (Altammami, Atwell, & Alsalka, 2020). Furthermore, the GitHub[3] repository of this corpus comes with a python script that can be used to extract any component of Hadith and its metadata.

5.2 Data Analysis

Before delving into the experiment, we give a brief overview of the two ontologies and the Hadith section-headings. Table 5 below shows the number of word tokens in each. By *word* we mean a token consisting of consecutive letters separated by space like the one in Figure 2, which is translated to an English sentence. Moreover, the table shows the number of distinct words which we call *word type*. Since Arabic is a highly inflected language, we extracted word lexemes using the Camel tool[4] (Obeid et al., 2020) to flatten the different forms of a word to its basic lexical unit. This will enable measuring the 'fit' as precisely as possible where a word like 'والإسلام'(and Islam) and 'للإسلام'(for Islam) collapse into the same lexeme 'إسلام'(Islam).

Table 5: Ontologies and Hadith headings vocabulary statistics.

	Word Tokens	Word Types	Lexeme
Bukhari Hadith Headings	39,959	6,110	2,913
Ontology A	41,782	978	746
Ontology B	103,603	2,009	1,370

So let them worship

فَلْيَعْبُدُوا۟

PRON   V   IMPV   CONJ

Figure 2: A word from Chapter 106 in the Quran.

5.3 Ontology Evaluation

As discussed earlier, we are following a data-driven approach to ontology evaluation. The idea is to find which ontology from the two candidates covers the Hadith better and which areas of the Hadith are not covered. Figure 3 below illustrates the steps taken to evaluate the ontologies. Then the following lines demonstrate the process taken to adopt this methodology.

---

[1] http://Quranytopics.appspot.com/
[2] http://www.Quranontology.com/
[3] https://github.com/ShathaTm/LK-Hadith-Corpus
[4] https://github.com/CAMeL-Lab/camel_tools

Figure 3: Ontology evaluation methodology.

1. Extract the concepts from ontology A (Abbas, 2009) and ontology B (Hakkoum & Raghay, 2016). Then extract the Hadith section-headings from LK Hadith corpus (Altammami et al., 2019).
2. Remove punctuations and diacritics, then tokenize the text by white space. This step produces three lists of words. One list contains the words of concepts from ontology A, the other list contains words of concepts from ontology B, and the final list contains words of Hadith section-headings.
3. To remove stop-words, we extracted all the Matn from the LK Hadith corpus to get the most frequent terms in Hadith statements. Then manually inspect the frequent terms to ensure they can be considered stop-words. These words were added to the list of Arabic stop-words list obtained from the NLTK, resulting in a list of 225 stop-words.
4. After removing stop-words from the three lists, we measured the overlap between the ontologies and Hadith section-headings lists. It turns out that only 38% of concepts in ontology A are in Hadith section-headings. However, we found words such as 'وتكذيبهم' which is translated to 'and their denial'. Hence, we decided to use the lexeme of words to measure the overlap instead of using the raw words.

6. Results

To measure the overlap between the two different ontologies and Hadith headings, the number of lexemes in the ontologies that also exist in Hadith headings lexemes is counted. In addition, the number of lexemes that exist in both ontologies is counted.

- 61.5% of the ontology A concept lexemes are present in the Hadith headings lexemes.
- 56.9% of ontology B lexemes are found in the Hadith headings lexemes.
- 100% of ontology A lexemes are present in ontology B lexemes. This implies that ontology A is incorporated within ontology B although it was not mentioned in their paper. However, the dedicated website for ontology B states that it is an ongoing research with constant updates.

The result of overlap is demonstrated in the diagram in Figure 4. Ontology B is 47% the size of the Hadith headings. While ontology A is 25% the size of Hadith headings as evident in the number of unique lexemes in each category as depicted in Table 5.
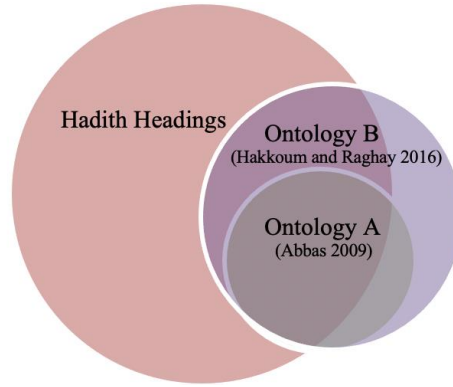
Figure 4: Overlap between Hadith section-headings and the Quran ontologies.

Since some parts of the ontologies do not have a match in the Hadith headings, we try to find synonyms of the ontology concepts that do not have a match as proposed by the methodology we are implementing (Brewster et al., 2004). Therefore, we used the Arabic Wordnet (Abouenour, Bouzoubaa, & Rosso, 2013) to extract synsets of the ontology concept lexemes that had no match in the Hadith headings lexeme list.

From 590 Quran lexemes, only 15% were found in the Arabic Wordnet. The resulting synsets are sensible as shown in the examples in Table 6. The Arabic Wordnet returns the synsets exactly as shown in the table, a pair of English and corresponding Arabic word or phrase (we did not do the translations). After scrutinizing the results, we conclude that the Arabic Wordnet is more appropriate for Modern Standard Arabic rather than Classical Arabic. In fact, important Islamic words were not found like زكاة - طهارة - صوم - فردوس (Paradise - fasting - purity - zakat). Therefore, we decided to not use the Arabic Wordnet.

Table 6: Synsets obtained from the Arabic Wordnet.

Term	Example of extracted synsets	Term	Example of extracted synsets
سقر	مكان_خيالي Imaginary_place سقر Hell	حشرة	حشرة Insect مفْصِلِيّ_الأرْجُل Arthropod
تطابق	استبدل Change إنتقل_من Change خفف_العقوبة Change تحضن_البيض Adjust	تحدى	تحدى Quarrel تصالح Argue هاجم Set_about غيظ Hurt فهم Accept
قرار	إخْتِيار Choice نشاط_بْشري Act طُرُق_قانُونِيّة Due_process	ندم	تعاسة Sadness حُزْن Sorrow تأْنِيب_الضمِير Compunction وَخْز_الضَمِير Compunction

7. Discussion

To understand the overlap illustrated in Figure 4, we enumerated some lexeme examples from each part as shown in Table 7. As one might predict, the lexemes that exist in all lists are the most important concepts. While the lexemes that exist in the ontologies are related to historical stories of prophets, messengers, and animals, with ontology B containing more details. This observation was statistically studied by (Sayoud, 2012) who remarked that several animal names are not cited in the Bukhari Hadith. For example, the name عجل (calf), is cited ten times in the Quran and is completely absent in the Bukhari Hadith.

Similarly, the lexemes in Hadith headings which were not covered in the ontologies are mostly about specific incidents or instructions that are only found in Hadith. For example, the complete instruction of Islamic washing that must be preformed before prayer is obtained from the Hadith because the Quran states the abstract obligation of the act without the details of how to perform it. Table 7 shows these words like مضمضة (mouth rinse).

Table 7: Examples of lexemes extracted from each part illustrated in Figure 4.

Overlap in Ontology A, B and Hadith headings		Overlap in Ontology A, B		Ontology B only		Overlap in Ontology B and Hadith headings		Hadith headings only	
مريم	قيامة	هابيل	ربوبية	عجل	تابوت	إبليس	بقر	مضمضة	لبن
الله	يوم	قابيل	وحداني	آدم	جالوت	شجرة	ذبح	مسح	أعرابي
جبرئيل	غزا	إلياس	تحدى	عصي	شريعة	نزول	قتيل	قدم	نفاس
إبراهيم	شيطان	عمران	جاحد	حواء	طائر	شخص	إسماعيل	وتر	صبي
يعقوب	ساجد	حواري	خلود	بعوض	أنعام	ألام	كعبة	تيمن	قاعد
مسجد	ملأك	نفاق	هداية	قدس	حصان	قثاء	مكة	حان	فرك
حج	آدم	عقيدة	طالح	عدس	ولادة	ثوم	مروة	قارئ	جنابة
نصراني	أكل	أخلاقي	فلاح	سينا	زكريا	بقل	صفا	غشي	إبل
يهودي	محرم	ذميم	سعادة	معتد	مهد	بصل	خنزير	توضأ	سمن
موت	إسرائيل	وعيد	تشبيه	بابل	طير	طور	رمضان	كعب	إمام
رجل	فرعون	فاسق	حشرة	إسحاق	قربان	رفع	فجر	نصيحة	رأس
توحيد	بحر	محيط	تذكير	إرجاع	غراب	قرد	قريش	شاة	إفشاء

8. Conclusion

In this paper, we investigate the appropriateness of using a Quran ontology as the basis for an Islamic ontology that covers the Hadith in addition to the Quran. Hence, the existing Quran ontologies were enumerated, discussed, then evaluated using a corpus-based evaluation approach. This evaluation was conducted by comparing the keywords extracted from the Bukhari Hadith book section-headings, and the Quran ontology concepts. The findings of this study can be understood through the presented visualization of overlaps in the Hadith keywords and the Quran ontology concepts. This allows the conclusion that ontology B (Hakkoum & Raghay, 2016) is the best candidate and could be used as a starting point for an Islamic ontology. Nevertheless, a large portion of Hadith is not covered by this ontology, which makes our next step is to explore methods that extend and incorporate the uncovered Hadith topics.

Acknowledgements

References

Abbas, N. (2009). Quran 'Search for a Concept' Tool and Website. *MSc Research Dissertation, Leeds University*, 1-170. https://www.researchgate.net/publication/318226723_Quran_'Search_for_a_Concept'_Tool_and_Website

Abouenour, L., Bouzoubaa, K., & Rosso, P. (2013). On the evaluation and improvement of Arabic WordNet coverage and usability. *Language Resources and Evaluation, 47*(3), 891-917. doi:10.1007/s10579-013-9237-0

Al-Yahya, M., Al-Khalifa, H., Bahanshal, A., Al-Odah, I., & Al-Helwah, N. (2010). An ontological model for representing semantic lexicons: An application on time nouns in the holy quran. *Arabian Journal for Science and Engineering, 35*(2 C), 21-35.

Alqahtani, M. M., & Atwell, E. (2018). Developing Bilingual Arabic-English Ontologies of Al-Quran. *2nd IEEE International Workshop on Arabic and Derived Script Analysis and Recognition, ASAR 2018*, 96-101. doi:10.1109/ASAR.2018.8480237

Alrehaili, S. M., & Atwell, E. (2018). Discovering Qur'anic Knowledge through AQD: Arabic Qur'anic Database, a Multiple Resources Annotation-level Search. *2nd IEEE International Workshop on Arabic and Derived Script Analysis and Recognition, ASAR 2018*, 102-107. doi:10.1109/ASAR.2018.8480361

Alromima, W., Moawad, I. F., Elgohary, R., & Aref, M. (2015). Ontology-based model for Arabic lexicons: An application of the Place Nouns in the Holy Quran. *11th International Computer Engineering Conference: Today Information Society What's Next?, ICENCO 2015*, 137-143. doi:10.1109/ICENCO.2015.7416338

Altammami, S., Atwell, E., & Alsalka, A. (2019). *The Arabic–English Parallel Corpus of Authentic Hadith.* Paper presented at the International Journal on Islamic Applications in Computer Science And Technology-IJASAT.

Altammami, S., Atwell, E., & Alsalka, A. (2020). *Constructing a Bilingual Hadith Corpus Using a Segmentation Tool.* Paper presented at the Proceedings of The 12th Language Resources and Evaluation Conference.

Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data Driven Ontology Evaluation. *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004*, 641-644.

Brown, J. A. (2017). *Hadith: Muhammad's legacy in the medieval and modern world*: Simon and Schuster.

Dukes, K. (2015). Statistical Parsing by Machine Learning from a Classical Arabic Treebank. PhD Thesis, Leeds University. Retrieved from http://etheses.whiterose.ac.uk/25746/

Gruber, T. R. (1991). The role of common ontology in achieving sharable, reusable knowledge bases. *Kr, 91*, 601-602.

Hakkoum, A., & Raghay, S. (2016). Ontological approach for semantic modeling and querying the Qur'an. *International Journal on Islamic Applications in Computer Science And Technology*, 37-37.

McDaniel, M., & Storey, V. C. (2019). Evaluating domain ontologies: Clarification, classification, and challenges. *ACM Computing Surveys, 52*(4). doi:10.1145/3329124

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., . . . Habash, N. (2020). *Camel tools: An open source python toolkit for Arabic natural language processing.* Paper presented at the Proceedings of The 12th Language Resources and Evaluation Conference.

Omar, A., Hyder, I., Iqbal, R., Murad, M. A. A., Mustapha, A., Mohd, N. S., & Mansoor, M. (2013). A Survey of Searching and Information Extraction on a Classical Text Using Ontology-based semantics modeling: A Case of Quran. *Life Science Journal, 10*(4), 1370-1377. doi:10.1017/CBO9781107415324.004

Raad, J., & Cruz, C. (2015). A Survey on Ontology Evaluation Methods. *HAL archives-ouvertes*.

Saad, S., Salim, N., Zainal, H., & Muda, Z. (2011). A process for building domain ontology: An experience in developing Solat ontology. *Proceedings of the International*

Conference on Electrical Engineering and Informatics, ICEEI(July). doi:10.1109/ICEEI.2011.6021572

Sayoud, H. (2012). Author discrimination between the holy Quran and Prophet's statements. *Literary and Linguistic Computing, 27*(4), 427-444. doi:10.1093/llc/fqs014

Sharaf, A. B. M., & Atwell, E. S. (2012a). QurAna: Corpus of the Quran annotated with Pronominal Anaphora. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 130-137.

Sharaf, A. B. M., & Atwell, E. S. (2012b). QurSim: A corpus for evaluation of relatedness in short texts. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 2295-2302.

Sherif, M. A., & Ngonga Ngomo, A.-C. (2009). Semantic Quran. *Semantic Web, 6*(4), 339-345.

Tashtoush, Y. M., Al-Soud, M. R., Abujazoh, R. M., & Al-Frehat, M. (2017). The noble quran Arabic ontology: Domain ontological model and evaluation of human and social relations. *8th International Conference on Information and Communication Systems, ICICS*, 40-45. doi:10.1109/IACS.2017.7921943

Ullah Khan, H., Muhammad Saqlain, S., Shoaib, M., & Sher, M. (2013). Ontology Based Semantic Search in Holy Quran. *International Journal of Future Computer and Communication, 2*(6), 570-575. doi:10.7763/ijfcc.2013.v2.229