

Growing Hadiths Ontology

Salah Alamri

seamari@uqu.edu.sa

Umm Al-Qura University, KSA

Abstract

The modern technological era has brought about the Semantic Web. Ontologies are essential to achieve the vision of the Semantic Web. Ontologies enable machines to understand data. The Arabic Language currently does not have a significant presence on the Web. To achieve a comparable level of Arabic access to other important languages, further work is needed to build Arabic ontologies. A goal is to design and create a robust Arabic ontology that represents the concepts from a large and significant subset of Arabic. We use a source of Hadiths (prophet saying and deeds) from *Riyadh As-Saliheen*. Preliminary results are very promising. **Keywords:** *Ontology, Hadiths Ontology, Growing Ontology, Ontology Extraction, Arabic Language.*

1. Introduction

The modern technological era has brought about the Semantic Web. The Semantic Web or Web 3.0 is intended to be an extension to the current Web (Web 2.0), enabling computers to communicate with humans better [1]. Its objective is to allow devices to understand complex concepts.

Ontologies are a gateway to achieving the Semantic Web vision. An ontology is “the formal, explicit specification of a shared conceptualization, which enables machines to comprehend meaning as well as share and reuse data [2].”

The Arabic language has an insufficient presence on the Web compared to other important world languages. Arabic is the sixth most spoken language after English, Mandarin,

Hindu, Spanish, and French [5]. Due to its widespread use, it is important to achieve a comparable level of Arabic access to other prevalent languages. In this paper, we discuss how to create a robust Arabic ontology, using the Sunnah (Hadiths) and the *Riyadh As-Saliheen* as references. We extract relationships between words to determine synonyms and antonyms. This process can also help further develop the Hadiths WordNet (HWN) [6]. We also extract relationships between concepts of Hadiths and other sources.

This paper has the following sections: Arabic Content, Ontology, Arabic Language and Arabic WordNet, Creating Ontologies, Growing Ontology, Ontology Extraction, and Conclusion and Future Work.

2. ARABIC CONTENT

Although there has been a desire to increase Arabic content on the Internet, a significant imbalance exists when compared to many other languages. A primary goal of Arabic content creators is to increase the amount of scientific research and other valuable topical content on the Web. Arabic is the language of the Quran and the Sunnah (Hadiths). The focus of this research is on the Sunnah Hadiths; the book *Riyadh As-Saliheen* was chosen due to its value and importance.

Riyadh As-Saliheen is composed of 435 Quran “verses,” containing 1,896 Hadiths, and it covers Islamic morals, acts of worship, and manners. The book conveys statements, actions, and tacit approvals of the Prophet Muhammad, Praise Be Unto Him (PBUH), and sometimes the Companion of Him [8]. Our process can be used with other Arabic books.

“Abu Hurairah (may Allah be pleased with him) reported: I heard the Messenger of Allah (PB upon him) saying: (I swear) by Allah, I seek Allah’s forgiveness and repent to Him more than seventy times a day [Bukhari] [9].”

”wean ’abiin hurayrat radi allah eanh qalan: smiet rasul allah salaa allh ealayh wsalam yaqul ”wallah ’iiniy l’astaghfr allaha, wa’atub ’iilyh, fi alyawmi, ’akthr min sabein mratan” rawah albukhari”

وَعَنْ أَبِي هُرَيْرَةَ رَضِيَ اللَّهُ عَنْهُ قَالَ: سَمِعْتُ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ يَقُولُ "وَاللَّهِ إِنِّي لَأَسْتَغْفِرُ اللَّهَ، وَأَتُوبُ إِلَيْهِ، فِي الْيَوْمِ، أَكْثَرَ مِنْ سَبْعِينَ مَرَّةً" رَوَاهُ الْبُخَارِيُّ

3. ONTOLOGY

An ontology “is simply the properties and relations between categories of a subject area.” An ontology is used to clarify the data in a way that enables machines to understand, share, and reuse data. It is a formal addition to the explicit specification of a shared conceptualization [2]. As an exact specification, it includes the types of concepts used and constraints on their use. As a shared conceptualization, it reflects consensual knowledge. Conceptualization refers to an abstract model of a phenomenon in the world by identifying relevant concepts of that phenomenon [4].

Ontologies can be constructed manually, semi-automatically, or automatically [7].

The ontology in Figure 1 describes a student being able to read a book and take a course but not manage the course, while a supervisor can manage the course and can supervise the student. Concepts and relationships between concepts are shown.

Ontologies can make explicit assumptions about systems, enable adaptability, and be reused. This flexibility reduces

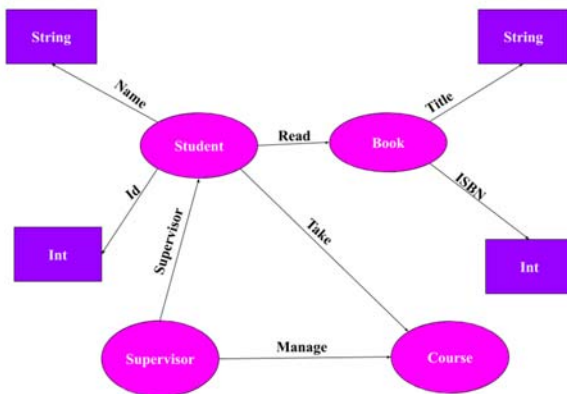


Figure 1: Example Ontology

development and operational costs, supports knowledge management and discovery, and can provide a basis for more adaptable systems [10][11].

4. ARABIC LANGUAGE AND ARABIC WORDNET (AWN)

There are many reasons why Arabic is important including Jahili Poetry and the Quran. The Quran is the Holy Book for all Muslims, which is in the origin of the Arabic language. The Prophet Muhammad (PBUH) taught the Islamic rules in the Arabic language, which is the primary reason behind its importance.

WordNet “is a lexical resource system on web.” Its design is inspired by the current linguistic theories of human lexical memory. In it, nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct lexical concept or sense. The Arabic WordNet (AWN) project has given the Arabic Natural Language Processing (NLP) community the first WordNet-compliant resource [3].

The Arabic language has multiple generations; the most important are the Classic Arabic Language and the Modern

Arabic Language. The classic Arabic Language is critical to understand major Islamic books, including the Quran and Hadiths, as well as Jahili poetry. The Modern Arabic Language is the most commonly used language in Arabic countries today. Thus, it is difficult for the Arabs to understand Islamic books written in Classic Arabic Language.

Most researchers in Arabic computational linguistics focus on Modern Arabic, while Classical Arabic, especially the Hadiths, is relatively unexplored though they are very important sources from the points of view of history, literature, and culture. Computational linguistics research of Hadiths is useful for both scholars and learners. Also, Hadiths help the Arabic language find the universal status as one of primary world languages [12].

5. CREATING ONTOLOGIES FROM ARABIC BOOK (Riyadh As-Saliheen Book)

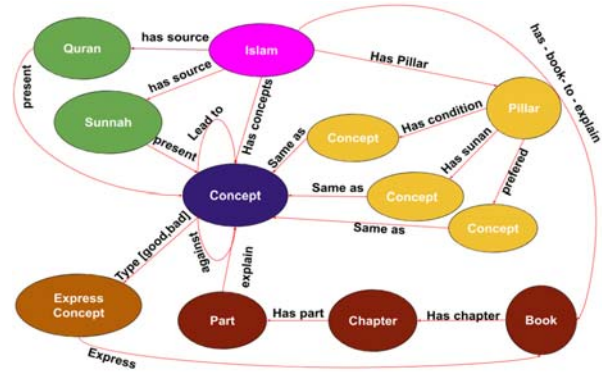


Figure 2: Islamic Ontology

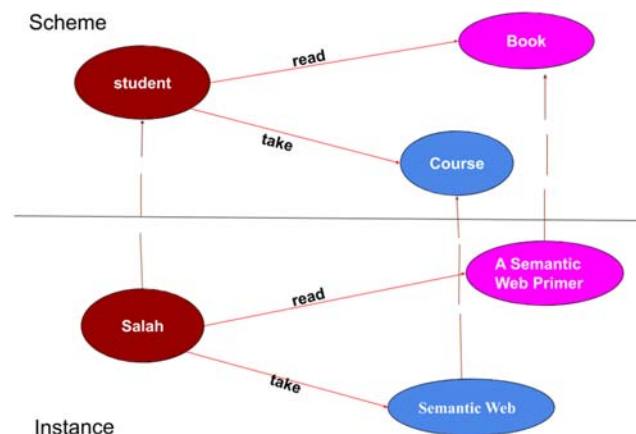


Figure 3: Ontology

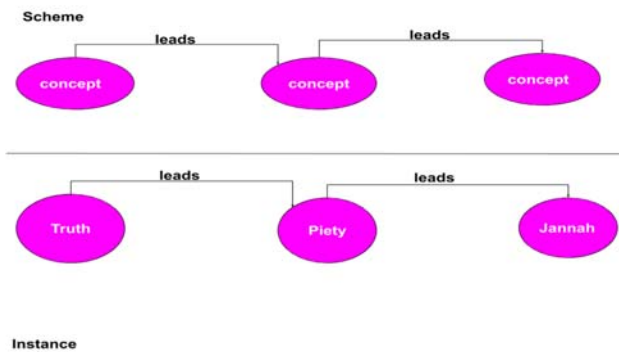


Figure 4: Hadith Ontology

We use the following four steps to build the ontology from the Arabic text (Riyadh as-Saliheen book). As we are validating our plans and goals, we are for now developing our ontology manually from texts. Our future plans include using semi-automatic methods to complete our goals.

First Select the domain and scope.

Second Extract the essential terms for the ontology from the domain.

Third Define the concepts.

Fourth Define the relationships between concepts.

A general Islamic ontology is shown in Figure 2. In Figure 3 and Figure 4 simple example ontologies are shown.

A. Select the Domain and the Scope of Ontology

The selected domain is the Islamic knowledge domain. We selected Hadiths from Riyadh as-Saliheen to study and investigate. For example, the Hadith: “Ibn Umar (Allah be pleased with both of them) reported Allah’s Messenger (may peace be upon him) as saying in connection with Ramadan:Do not fast till you the new moon, and do not breakfast till you it; but if the weather is cloudy calculate about it ”[9].

Another Hadith is “The Messenger of Allah (PUH) said, ”He who goes forth in search of knowledge is considered as struggling in the Cause of Allah until he returns.”[9].

B. Extracting the Essential Terms in the Ontology

To extract the essential terms, we need to remove the frequent unnecessary stop words such as articles and other words which are not important as shown Figure 5.

Hadiths	Terms	Hadiths(Aabic)
Messenger as saying in connection with Ramadan: Do not fast till you see the new moon, and do not break fast till you see it; but if the weather is cloudy calculate about it. (The Hadith is intended here to people and the new moon here means the start of the new month and Ramadan Complete mean the month Finished).	People, Ramadan, Ramadan Complete, Do not fast, Till see, New moon, Do not break fast, Till see, Weather, Cloudy calculate	عن النبي صلى الله عليه وسلم أنه ذكر رمضان فقال لا تصوموا حتى تروا الهلال ولا تطروا حتى تروا الهلال ولا تحسبوا عظيم فاغفروا له
“He who goes forth in search of knowledge is considered as struggling in the Cause of Allah until he returns.”	People, Go forth, Search, Knowledge, Consider struggle cause of, Allah, Until, He, Return	قال رسول الله صلى الله عليه وسلم: "من خرج في طلب العلم، فهو في سبيل الله حتى يرجع" The pronoun here in this sentence means people.

Figure 5: Extracting the Essential Terms from Hadiths

C. Defining the Concepts

This step includes selecting the noun words to represent the classes and the nouns to represent instances within a class. Generic nouns are concepts, for example, student, professor, and course. Nouns representing individuals or individual things are instances or entities such as Salah, Austin, Semantic Web. Please see Table 1 and Figure 12 in section 7

TABLE 1: Defining the Instances

Concept	Instance
student	Salah
professor	Austin
course	Semantic web

D. Defining the Relationships between Concepts

This essential step is when relationships between extracted concepts are determined. This is supported by the use of patterns such as Noun – Verb – Noun as is shown in Table 2 and Figure 6 to Figure 8.

TABLE 2: Relationships between Concepts

Domain	Predicate	Range
class	forbade about	class
class	Don’t meet	class
class	Heard about	class
class	Don’t enter	class
class	Don’t leave	class
class	Till arrive	class
class	weak up	class
class	Till wash	class
class	commanded	class
class	leads to	class

class	keep	class
class	register	class
class	perform to	class
class	rinse to	class
class	enroll	class
class	incur loss	class
class	get wind outbreak	class
class	breaks out	class
class	do not enter	class
class	do not leave	class
class	is in	class
class	are in	class
class	hasType	class
class	IsTypeof	class
class	hasPartOf	class

Hadiths	Patterns	Result	Hadiths(Arabic)
“Do not meet the merchandise till they arrive in the market.” (The Hadith is intended here to people).	Noun-verb-noun-verb-noun	People- do not meet-merchandise- till arrive -the market	”لا تتلقوا السلع حتى يهبط بها إلى الأسواق“

Figure 8: Relationships between Concepts

6. GROWING ONTOLOGY

We collected some sources from Hadiths to grow our ontology and showed how to process them. Now we take a whole (single) Hadith or a part of a (single) Hadith, and we build an ontology scheme, which is the ontology structure consisting of the concepts, entities, and relations which we have collected and determined. We do this for the relevant Hadiths mentioned by the Prophet about truth, as shown in Figure 9.

Hadiths	Patterns	Result	Hadiths(Arabic)
“If you get wind of the outbreak of plague in a land, do not enter it; and if it breaks out in a land in which you are, do not leave it.” (The Hadith is intended here to people).	Noun-Verb- (Noun Noun)- Verb- (Noun noun)	People- get wind outbreak - Plague Plague - is in - Land- 1 People -do not enter- Land-1 Plague - breaks out - Land -2 People -are in- Land -2 People -do not leave- Land-2	”إذا سَمِعْتُمْ بِالطَّاعُونِ بِأَرْضٍ فَلَا تَدْخُلُوهَا، وَإِذَا وَقَعَ بِأَرْضِكُمْ وَأَنْتُمْ بِهَا فَلَا تَخْرُجُوا مِنْهَا“
“There are two blessings in which many people incur loss. (They are) healthy and free time (for doing good)”	(Noun Noun)- verb - Noun	Many people-incur loss-Health Many people-incur loss- free time	”نِعْمَتَانِ مَغْفُورٌ فِيهِمَا كَثِيرٌ مِنَ النَّاسِ، الصُّحَّةُ وَالْفَرَاغُ.“

Figure 6: Relationships between Concepts

Hadiths	Patterns	Result	Hadiths(Arabic)
“Truth leads to piety and piety leads to Jannah. A Person keeps talking the truth until he is registered with Allah as truthful. Falseness leads to vice and vice leads to the fire (Hell), and a person keeps on telling lies until he is enrolled as a liar.”	Noun-verb - noun	Truth -leads to-piety. piety- leads to Jannah. Person- keep- truth. he register truthful. Falseness - leads to- vice. vice -leads to- the fire. person- keep- lies. he enroll liar.	إن الصدق يهدي إلى البر، وإن البر يهدي إلى الجنة، وإن الرجل ليصدق حتى يكتب عند الله صديقاً، وإن الكذب يهدي إلى الفجور، وإن الفجور يهدي إلى النار، وإن الرجل ليكذب حتى يكتب عند الله كاذباً
“If you perform ablution, then rinse your mouth.” (The Hadith is intended here to person).	Noun - verb-noun-verb-noun	Person-perform to- ablution-rinse to - mouth	إذا توضأ فليضمض

Figure 7: Relationships between Concepts

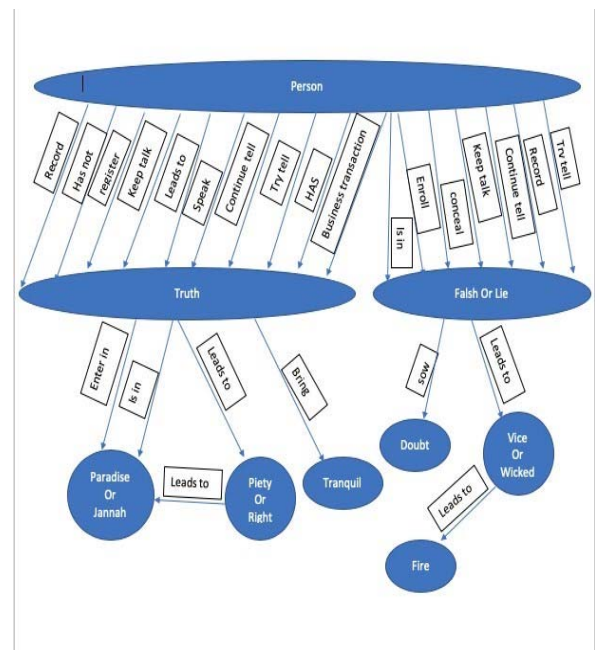


Figure 9: Grow Ontology

7. ONTOLOGY EXTRACTION

To build ontologies from Arabic sources we can use the methodology described in Figure 10.

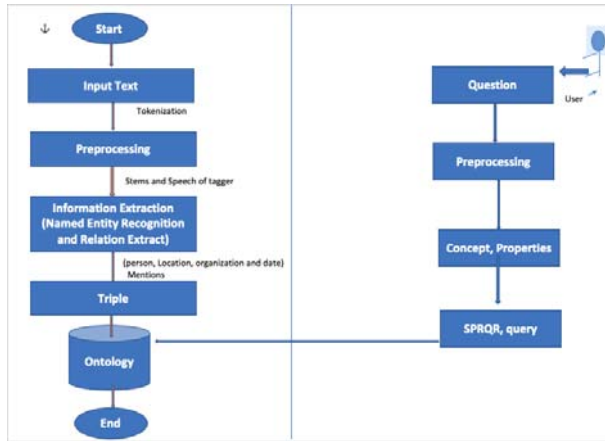


Figure 10: Build Ontology from Arabic

First: Input Text.

In our research we have used Riyadh as-Saliheen as a source.

“The person who goes about calumnies will never enter Jannah”.

لا يدخل الجنة نمام

Second: Preprocessing.

This stage includes the text preparation, e.g., Tokenization, Part-of-Speech tagging, and Stemming, etc.)[13].

A. The Tokenizer divides the text into tokens:

Ex. (calumnies- -never nter Jannah)).

(لا يدخل- الجنة- نمام)

B. Part-of-Speech adds part-of-speech information to tokens:

EX. (calumnies (Noun) -never _enter(verb)- Jannah(Noun)).
(اسم) لا يدخل(فعل)- الجنة(اسم)- نمام.)

C. Here the stemming is root extraction:

Ex. Enter- Jannah – Calumny.

(لا يدخل (دخل) – الجنة (جنة) – نمام (نم))

D. Finally, remove stop words, Latin words, single-letter words, prepositions, and conjunctions.

Ex. Never enter- Jannah – Calumny.

لا يدخل – الجنة- نمام

Third: Information Extraction.

Information extraction contains the identification of Named Entity Recognition (NER) Relationships between

Entities. An example of Named Entity Recognition (NER) is shown in TABLE 3.

TABLE 3: Named Entity Recognition (NER)

NER	Arabic Word	English Word
Location	الجنة	Jannah
Person	نمام	Calumny

The relationships between these Entities is depicted in Figure 11. and Figure 12..



Figure 11: Relationships between Entities



Figure 12: Relationships between Entities

Fourth: Finally, the ontology is assembled.

8. CONCLUSION AND FUTURE WORK

The Arabic Language has an insufficient presence on the Web. Therefore, we are trying to increase the content of Arabic to achieve an Arabic Semantic Web more comparable with other important languages such as English and Chinese. We built Arabic ontologies to realize the Semantic Web vision. The goal is to create a robust Arabic ontology that represents the concepts from a large and significant subset of Arabic.

References

[1] Berners-Lee, T., Hendler, J. and Lassila, O. *The Semantic Web*. Scientific American, 284(5), 34-43 (2001).

[2] Gruber, T. *Toward principles for the design of ontologies used for knowledge sharing*. International Journal of Human-Computer Studies, 43(5), 907-928 (1995).

[3] Al-Zoghby, A., Ahmed, S., Hamza, T. *Arabic Semantic Web Applications. A Survey*. Journal of Emerging Technologies in Web Intelligence 5(1). (2013).

- [4] Studer, R., Benjamins, V., and Fensel, D. *Knowledge Engineering: Principles and Methods*. Data Knowledge Engineering 25(1), 161-197 (1998). doi=10.1.1.110.8406.
- [5] <https://www.visualcapitalist.com/100-most-spoken-languages/>. Accessed: May 30, 2021.
- [6] Alkhatib, M., Monem, A., Shaalan, K. A Rich Arabic WordNet Resource for Al-Hadith Al-Shareef. *Procedia Computer Science*, 117, 101-110. (2017)
- [7] Hazman, M., El-Beltagy, S., and Rafea, A. *A Survey of Ontology Learning Approaches*. *International Journal of Computer Applications* 22(9). (2011).
- [8] Alosaimy, A., Atwell, E. *Sunnah Arabic Corpus: Design and Methodology*. Proceedings of the 5th International Conference on Islamic Applications in Computer Science and Technologies. IMAN (2017).
- [9] Al-Nawawi, M. *Riyadh As-Saliheen*.
- [10] Anjana, G. and Singh, J. *Quality metrics for data warehouse multidimensional models with focus on dimension hierarchy sharing*. *Advances in Intelligent Informatics*, 429-443. Springer International Publishing, (2015). doi.10.1007/s13198-017-0641-5.
- [11] Ahmad, K., Hirzalla, N., Samhan, A. and Alfayoumi, M. *Towards an ontology for software product quality attributes*. *Internet and Web Applications and Services*, 2009. ICIW 09. Fourth International Conference on Internet and Web Applications and Services, 200-204. IEEE, (2009).
- [12] Alkhatib, M., Monem, A., Shaalan, K. *A Rich Arabic WordNet Resource for Al-Hadith Al-Shareef*. *Procedia Computer Science*, 117, 101-110. (2017)
- [13] Maynard, D., Li, Y., and Peters, W. *NLP Techniques for Term Extraction and Ontology Population*. Dept. of Computer Science, University of Sheffield, UK, (2008)



Salah Alamri received his B.S. degree in Computer Science from King Abdulaziz University, Saudi Arabia in 2010, and received M.S. degree in Computer Science from Kent State University, USA in 2014. He received the PhD degree in Computer Science in August 2020 from Kent State University, USA. He is currently work as an assistant

professor, Computer Science department in the Faculty of Computing At Al-Qunfudhah branch at Umm Al-Qura University.