

Arabic Ontology for Hadith texts - A survey

Mariam Muhammed^{*1}, Shahira Azab^{**2}, Nesrin Ali^{*3}, Mervat Ghieth^{**4}

** Department of Information Systems and Technology, Faculty of Graduate Studies for Statistical Research (FGSSR), Cairo University, Egypt.*

¹eng.maryamadel@cu.edu.eg

³Nesrineali79@cu.edu.eg

***Department of Computer Science, FGSSR, Cairo University, Egypt.*

²shahiraazazy@cu.edu.eg

⁴mervat_gheith@yahoo.com

Abstract: Hadith texts play a significant role in Islamic scholarship, providing guidance for Muslims in understanding and practicing Islam. This paper explores the importance of Hadith texts, their categorization based on reliability, and the structure of a typical Hadith. It discusses the use of software tools and ontologies in analysing Hadith texts. The paper also examines the Arabic language, particularly its complexities and the challenges it poses in Natural Language Processing tasks. Arabic is classified into Classical, Modern, and dialectal forms. Arabic has many challenges that affect the building of ontologies; for instance, the lack of linguistic resources and complex morphology. The paper also explores the development of ontologies in Islamic studies, specifically for Hadith, highlighting their unique characteristics and applications. Several research works related to Hadith ontology and its potential applications are discussed, including the creation of Al-Hadith WordNet, computational analysis of Hadith texts, and the development of ontology-based systems for authentication and retrieval of Hadith information.

Key words: Al-Hadith, Hadith Ontology, Islamic ontologies, Arabic Language.

1 INTRODUCTION

In Arabic, the word "hadith" has different meanings like communication, story, or conversation. This can include religious or non-religious, historical or current topics. According to theologians, a hadith specifically refers to anything that was transmitted based on the authority of Prophet Muhammad (Peace be Upon Him). This includes his actions, words, silent approvals, or descriptions of his physical appearance. The Holy Qur'an tells Muslims to follow what Prophet Muhammad did, so his companions started to focus on following his ways, called the sunna. These ways were recorded in hadiths, which are stories about what the Prophet said and did. For Islamic legislation, the Holy Qur'an and hadith are the two main sources. In traditional Muslim schools of jurisprudence, hadith is very important for understanding the Qur'an and all jurisprudence matters.

The importance of hadith texts lies in their role as a guide for Muslims in understanding and practicing Islam. These texts contain valuable teachings, actions, and sayings of Prophet Muhammad (Peace be Upon Him), providing practical examples of how to follow the teachings of the Holy Qur'an in daily life. Hadiths help clarify ambiguous or complex aspects of Islamic teachings, offering guidance on matters such as worship, morality, social conduct, and jurisprudence. They serve as a source of inspiration, wisdom, and ethical guidance for Muslims worldwide, helping them to live in accordance with the principles of Islam and to uphold the values of compassion, justice, and righteousness [1].

The Holy Qur'an and hadith are very important in old Arabic writing. People still want to read old Arabic texts for many reasons. In Arabic culture today, old texts are more valued than in some Western countries. Many websites offer these texts to anyone interested. The hadith text is especially big in old Arabic writing. An organization called the European Language Resource Association recognizes how important the hadith is and has given it a reference number: ELRA-WC0134 [2].

While the Qur'an is believed to be the pure word of God, the hadith collection includes many different texts with different levels of reliability, ranging from very reliable to not reliable at all. Generally, hadiths are divided into four categories based on how trustworthy they are: sahih, which means highly accurate; hasan, meaning they're pretty reliable; dha'if, indicating they're not very reliable; and maudu', which means they're fake or fabrication [3].

Every hadith has two parts: the sanad (or isnad) and the matn. The matn is the actual text of the hadith, while the sanad tells us how the matn was passed down. The sanad, also called the narration chain, lists narrators in reverse order, each one saying who they heard the hadith from, until reaching the original narrator of the matn, followed by the text of the matn itself [4]. The Sanad is like a list of people who passed down a Hadith, with each person saying where they heard it from, until it goes back to the main narrator of the story, called the Matan. The Matan is the actual content of the Hadith. The Sanad is really important because it helps scholars figure out if a Hadith is real or not. They look at the chain of people who passed it down to decide if they trust it or not. Fig.1 displayed the structure of a hadith using an example taken from the book "Riyad al-Salihin."

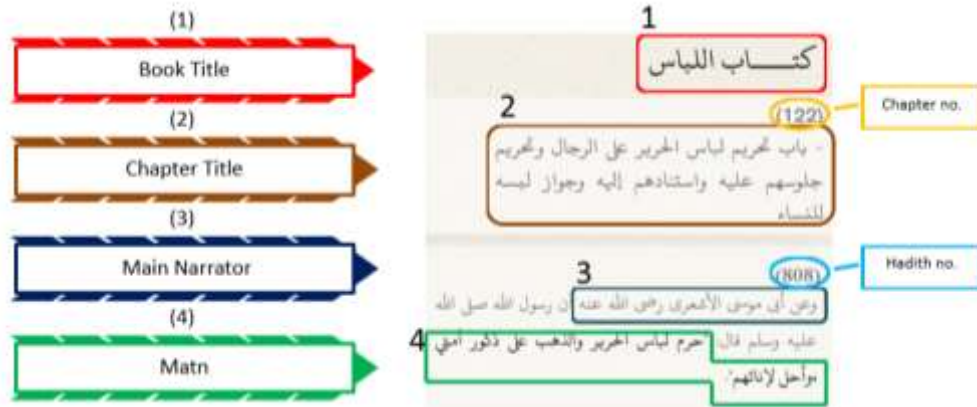


Figure 1 A sample Hadith taken from "Riyad al-Salihin" book.

Software tools like electronic Al-Hadith encyclopedias and certain Hadith websites are used to assess specific Isnads. More recently, tools like ontologies related to the semantic web have been employed in this process. An ontology is a structured description of concepts in a particular domain, including the properties and attributes of each concept. It serves as a foundation for various applications, such as Information Retrieval systems and Decision-Support Systems. When combined with instances of concepts, ontologies form a knowledge base [5].

On the other hand, various researchers worldwide, including scholars in Natural Language Processing and specific research groups, concentrate on studying the Matn. For instance, authors in [6] focuses on classifying Al-Hadith into various chapters using distinct techniques. Additionally, in [5], a text mining tool was created to search queries within an Al-Hadith dataset, offering a list of relevant Hadiths ranked by their similarity.

In this paper, we have gathered some of research studies that focused on Al-Hadith.

The remainder of this paper is organized as follows: Section 2 provides an overview of the Arabic language. Section 3 introduces an ontology in Islamic studies, specifically for Hadith, highlighting the major differences between Hadith ontology and ontologies for other domains, along with a brief introduction of WordNet. Section 4 presents some works related to the Hadith domain. Section 5 discusses previous works on using ontology for Hadith. Finally, in Section 6, we present conclusions and future work.

2 ARABIC LANGUAGE OVERVIEW

Experts today divide Arabic, a type of Semitic language, into three main groups: Classical, Modern, and dialectal Arabic. Classical Arabic (CA) refers to the language spoken by people from the Arabian Peninsula around the sixth century Common Era (CE), back when Islam was emerging. Standard Arabic is what earlier Arab language experts thought of as "Classical," while Modern Standard Arabic (MSA) is how people today use Standard Arabic in their daily lives [7].

The Arabic language, with its rich history and diverse dialects, presents both challenges and opportunities in the realm of building ontologies, particularly when it comes to classical Arabic. Ontologies, which serve as structured representations of knowledge, are crucial in various fields such as artificial intelligence, semantic web, and information retrieval. However, constructing ontologies in Arabic poses specific difficulties due to linguistic complexities and the evolution of the language over time.

One of the primary challenges in building ontologies in classical Arabic is the standard form of the language. While classical Arabic serves as the language of the Quran and classical literature, its usage has evolved over centuries, leading to discrepancies between classical and modern Arabic. This evolution introduces ambiguity and inconsistency in interpreting classical texts, making it challenging to extract precise semantic information for ontology development. Another significant issue is the lack of comprehensive linguistic resources and tools tailored specifically for ontology construction in Arabic. Unlike languages such as English, which have well-established linguistic resources and ontological frameworks, Arabic lacks comparable resources, hindering the development of robust ontologies. Limited access to annotated corpora, lexicons, and semantic parsers impedes the accurate representation of Arabic semantics within ontological structures. Moreover, the complex morphology and syntax of Arabic pose computational challenges for automated ontology extraction and development. Arabic morphology, characterized by root-and-pattern morphology and extensive inflectional processes, requires sophisticated natural language processing (NLP) techniques to handle effectively.

Parsing Arabic sentences, identifying semantic relationships, and disambiguating word meanings demand advanced NLP algorithms tailored to the intricacies of the Arabic language.

To address these challenges, it requires collaborative efforts from linguists, computer scientists, and domain experts to develop specialized tools and methodologies for ontology development in Arabic. Collaborative initiatives aimed at compiling comprehensive linguistic resources, standardizing vocabularies, and refining NLP algorithms can facilitate the creation of high-quality Arabic ontologies. Additionally, incorporating domain-specific knowledge and expert insights is crucial for ensuring the relevance and accuracy of ontologies in diverse application domains.

3 ARABIC ONTOLOGY

Ontologies are used to show what we know about a certain area or topic, like the world or a specific field. They're important in Artificial Intelligence, the Semantic Web, and other areas like Software Engineering and Biomedical Informatics. In languages like English and French, there are many well-developed ontologies. However, there aren't as many in Arabic, especially for Islamic studies [8]. As Islamic sciences progressed, it became clear that there was a need to create ontologies in various fields of Islamic studies.

A. Islamic Ontologies vs. other Ontologies

Hadith ontology as one of the Islamic ontologies, like other ontologies in various domains, serves as a structured representation of knowledge within its specific domain. However, the nature and characteristics of Hadith ontology distinguish it from other ontologies in several key ways. Firstly, the domain of Hadith ontology pertains to Islamic teachings and traditions, particularly the sayings and actions of the Prophet Muhammad as recorded by his companions. This domain specificity imbues Hadith ontology with unique religious and cultural significance, shaping its content and applications. Secondly, the structure of Hadith ontology reflects the hierarchical and relational nature of Islamic scripture. Hadiths are categorized based on their authenticity, chains of narrators, and thematic content, necessitating a complex ontology structure that accounts for these nuances. This structural complexity differs from other ontologies, which may prioritize different organizational schemes based on the requirements of their respective domains. Moreover, the development and validation of Hadith ontology involve expertise from Islamic scholarship, textual analysis, and computational linguistics, among other disciplines. This interdisciplinary approach ensures that the ontology accurately represents Islamic teachings while also adhering to standards of ontology engineering. Furthermore, the applications of Hadith ontology often revolve around Islamic scholarship, education, and information retrieval systems tailored to the needs of Muslim communities worldwide. These applications cater to specific religious and cultural contexts, facilitating access to authentic Hadith sources and promoting a deeper understanding of Islamic teachings.

In contrast, other ontologies span diverse domains such as healthcare, finance, and engineering, each with its own set of stakeholders, standards, and applications. While these ontologies may share common principles of ontology design and implementation, their content, scope, and objectives vary significantly based on the needs and requirements of their respective domains. Table 1 summarizes the key difference between Hadith ontology and other ontologies in different domains.

TABLE 1

Key Differences Between Hadith Ontology and Other Ontologies in Different Domains.

Aspect	Hadith Ontology	Other Ontologies
Domain	Islamic teachings and traditions	Varied domains such as healthcare, finance, engineering, etc.
Content	Sayings and actions of Prophet Muhammad, categorized based on authenticity, narrators, and thematic content	Content specific to the domain, structured based on relevant concepts, entities, and relationships (other ontologies)
Applications	Primarily focused on Islamic scholarship, education, and information retrieval systems for Muslim communities	Diverse applications tailored to specific domains, such as clinical decision support systems, financial risk management, etc.
Purpose	Facilitates access to authentic Hadith sources, promotes understanding of Islamic teachings	Facilitates knowledge organization, reasoning, and application within specific domains

B. WordNet

WordNet, created by A. Kilgarriff in [9], is a popular tool for understanding words in languages. It groups words like nouns, verbs, adjectives, and adverbs into groups of similar meanings called synsets. Each synset shows one idea, and they're linked by how they relate to each other, like how some words are more general or specific than others. WordNet has been used to make many other language tools in different languages.

Amine Arabic Word Net is a special tool made for the Arabic language. It's free to use and is based on Princeton WordNet. It can also connect with other WordNets for various languages, making it possible to translate between many different languages [10].

Alkhatib et al. [11] proposed a new approach to create Al-Hadith WordNet, a useful language tool specifically tailored for hadith texts. Unlike the Modern Standard Arabic (MSA) WordNet, which exists, there wasn't a similar resource for the Classical Arabic found in hadith texts. To fill this gap, the authors developed Al-Hadith WordNet by carefully analyzing and preparing hadiths from Sahih of Bukhari and Sahih of Muslim. This resource allows users to access words, their synonyms, and the relationships between them based on lexical and semantic connections. The effectiveness of Al-Hadith WordNet was tested by applying it to about 8500 sets of synonyms and then using it to categorize hadith texts into their respective chapters. The result was the accurate classification of 2671 hadiths into 24 chapters, with an impressive average F-measure of 94.5%. The approach of Al-Hadith WordNet is illustrated in Fig. 2.

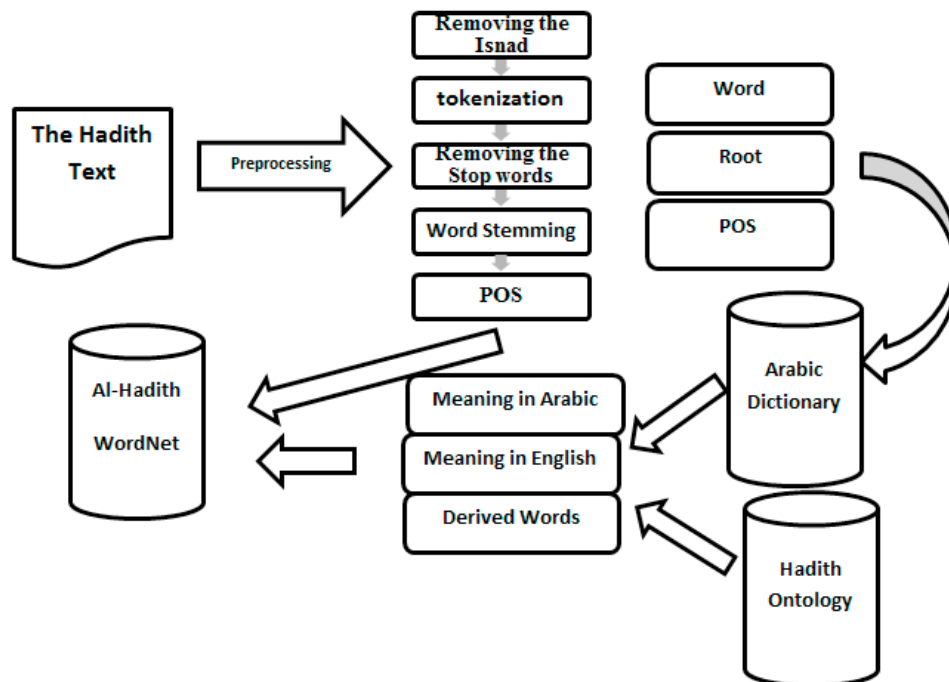


Figure 2 A Framework for building Al-Hadith WordNet [11]

4 HADITH CORPUS

Hadiths consist of the spoken words and actions of Prophet Muhammad (peace be upon him). In traditional Muslim belief, Hadiths are essential for understanding the Quran and play a crucial role in legal affairs [12].

According to Altammami et al. [13], every hadith represents an individual statement or action of the Prophet Muhammad, which was recorded, gathered, and compiled into books. Scholars then categorized these hadiths based on their deep understanding and extensive knowledge of the subject matter.

As mentioned before, the Hadith is divided into two parts: the main narrative, known as matan, and the sequence of narrators through whom the narration has been passed, traditionally known as Isnad. Isnad is a list of narrators in chronological order, with each person mentioning who they heard the hadith from, ultimately leading back to the primary narrator of the Matan, followed by the Matan itself [14].

According to [15], the term "al-Jadid" or "the new" is used to refer to the Hadith in contrast to "al-Qadim" or "the ancient". Hadith is typically understood as teachings derived from the Prophet (peace be upon him). Although Hadith is also called Sunnah, there's no difference between the two in essence. However, Hadith specifically refers to the texts found in collections, while Sunnah encompasses the actions and practices of Muslims in their daily lives.

According to Ibn Taymiyyah, both hadith and sunnah carry the same significance, encompassing everything that occurred involving the Prophet after receiving prophethood, including his words, actions, and decisions [16].

Many scholars have written explanations of Hadith. They describe and make clear the Hadith, especially for people who don't speak Arabic well. Without these explanations, the true meaning of the Hadith might not be understood correctly, which could lead to mistakes in understanding it [17]. According to [18], understanding Hadith isn't just about using a dictionary and a book of Hadith as references. This is because there are many books with Hadith in them. So, Takhrij al-hadith seems to be a way to find out where a Hadith originally came from.

A. Markup Tags in a Hadith Corpus

Markup tags in a Hadith corpus are used to encode structural and semantic information within the text of the Hadiths. They serve to annotate different elements of the text, making it easier to analyze, search, and manipulate the corpus [19]. Common markup tags found in a Hadith corpus may include:

- **Hadith Text Tags:** These tags encapsulate the actual text of the Hadith itself, often demarcating the beginning and end of each Hadith to differentiate them from one another.
- **Narrator Tags:** These tags enclose the names of the narrators (Rijal) in the chain of transmission (Isnad) of the Hadith. They may include information about the narrator's full name, kunya (honorific), title, etc.
- **Chain of Transmission Tags:** Tags used to mark the entire chain of narrators, indicating the sequence of transmission from the Prophet Muhammad (peace be upon him) to the compiler of the Hadith collection.
- **Attribution Tags:** Tags indicating the attribution of the Hadith to a particular source or compilation, such as Sahih al-Bukhari, Sahih Muslim, etc.
- **Isnad Tags:** These tags may specifically mark the Isnad portion of the Hadith, indicating the chain of narrators leading back to the Prophet Muhammad (peace be upon him).
- **Matn Tags:** Similarly, there may be tags specifically marking the Matn portion of the Hadith, which comprises the actual content or text of the Hadith itself.
- **Metadata Tags:** These tags enclose metadata about the Hadith, such as its unique identifier, date of compilation, category or subject matter, and any additional notes or comments.
- **Markup for Authenticity:** Some Hadith corpora may include markup indicating the authenticity rating of each Hadith, such as "Sahih" (authentic), "Hasan" (good), "Da'if" (weak), etc.
- **Cross-Reference Tags:** Tags used for cross-referencing related Hadiths or other texts within the corpus, facilitating navigation and analysis.

These markup tags help in structuring and organizing the Hadith corpus, enabling researchers, scholars, and practitioners to extract and analyze information efficiently. The specific markup conventions may vary depending on the standards adopted by the compilers or publishers of the corpus. Table 2 shows a list of Hadith Datasets.

B. Tools for Linguistic Processing of Hadith Texts

Several natural language processing (NLP) tools and libraries can be utilized for tokenization, lemmatization, stemming, and other linguistic processing tasks on Hadith texts [20]. Some commonly used tools include:

- **Farasa Toolkit [21]:** Farasa Toolkit is an Arabic natural language processing toolkit developed by the Qatar Computing Research Institute. It offers various functionalities for Arabic text processing, including tokenization, stemming, named entity recognition, part-of-speech tagging, and morphological analysis. Farasa Toolkit is particularly known for its accuracy and efficiency in handling Arabic text [22].
- **Camel Tools [23]:** CAMEL Tools is an open-source Python toolkit. It offers a suite of tools for various NLP tasks, including pre-processing, morphological modeling, dialect identification, named entity recognition, and sentiment analysis. It provides utilities that enable researchers and developers to work with Arabic text effectively, facilitating tasks such as tokenization, stemming, and part-of-speech tagging [24].

TABLE 2
List of Available Hadith Datasets.

Corpus Name	Description	Tasks	Availability
LK-Hadith-Corpus¹ [25]	A parallel corpus containing Islamic Hadith in two languages.	machine-translation	Free
The Nine Books Of Arabic Hadith²	Hadith texts with diacritical marks and without diacritical marks.	text-classification text-Similarity	Free
Quran Hadith Datasets³	The datasets contain pairs of Quranic verses with both related and unrelated Hadiths or other Quranic verses. They include Classical Arabic text alongside English translations of verses and teachings.	semantic-similarity	Free
Sanadset 650K: Data on Hadith Narrators [26]	A total of 650,986 records have been gathered from 926 historical Arabic Hadith books.	Hadith authentication machine-translation topic-classification named-entity- recognition Question-answering information-retrieval natural-language-inference	Free
Hadith Narrators Dataset (+24K)⁴	Over 24,000 narrators and scholars of Hadith are represented, spanning a period of more than 500 years.	Hadith authentication topic-classification named-entity- recognition Question-answering information-retrieval Social Network Analysis	Free
Hadith Project⁵	The collection comprises 7,658 Arabic hadiths along with the names of 1,755 transmitters.	Hadith authentication Information-retrieval Named-Entity-Recognition Social Network Analysis	Free

- **Stanford CoreNLP [27]:** Stanford CoreNLP is a natural language processing toolkit developed by the Stanford NLP Group. It provides various NLP functionalities, including tokenization, sentence splitting, part-of-speech tagging, lemmatization, named entity recognition (NER), sentiment analysis, dependency parsing, coreference resolution, and more. Stanford CoreNLP supports multiple languages, including Arabic.
- **NLTK (Natural Language Toolkit) [28]:** NLTK is a popular Python library for NLP tasks. It provides functionalities for tokenization, stemming, lemmatization, part-of-speech tagging, and more. Users can employ NLTK for preprocessing Hadith texts before further analysis.

¹ <https://github.com/ShathaTm/LK-Hadith-Corpus>

² <https://github.com/abdelrahmaan/Hadith-Data-Sets>

³ https://github.com/ShathaTm/Quran_Hadith_Datasets

⁴ <https://www.kaggle.com/datasets/fahd09/hadith-narrators>

⁵ <https://www.kaggle.com/zusmani/hadithsahibukhari?select=Sahih+Bukhari+Transmitter+Names.txt>

- **SpaCy** [29]: spaCy is another powerful NLP library for Python. It offers efficient tokenization, lemmatization, part-of-speech tagging, dependency parsing, and named entity recognition. spaCy's pre-trained models can be used to process Hadith texts effectively.
- **Stanza** [30]: Stanza is a Python NLP library developed by the Stanford NLP Group. It provides tokenization, lemmatization, part-of-speech tagging, dependency parsing, and named entity recognition capabilities. Stanza's models support multiple languages, including Arabic, making it suitable for processing Hadith texts.
- **Gensim** [31]: Gensim is a Python library primarily used for topic modeling and document similarity analysis. It also offers utilities for tokenization, stemming, and lemmatization. Gensim's preprocessing functionalities can be applied to Hadith texts to prepare them for further analysis.
- **Arabic NLP Libraries**: There are specialized NLP libraries specifically designed for processing Arabic texts, such as Buckwalter Arabic Morphological Analyzer (BAMA) [32], and MADAMIRA [33][34]. These libraries offer tokenization, stemming, lemmatization, and other linguistic processing tools tailored to the unique characteristics of Arabic language.

These tools and libraries can aid in preparing Hadith texts for various NLP tasks, enabling researchers to extract meaningful insights and analyze the corpus effectively. The choice of tool depends on factors such as the specific requirements of the analysis and the availability of resources. Table 3 presents a comparison of natural language processing tools and libraries for Arabic text processing

TABLE 3
Comparison of Natural Language Processing Tools and Libraries for Arabic Text Processing.

	Camel Tools	Farasa	Stanford CoreNLP	MADAMIRA
Language	Arabic Specific	Arabic Specific	Multilingual	Arabic Specific
Developed Language	Python	Java	Java with Python bindings	Java
Exposed Pre-processing	√	-	-	-
Morphological Disambiguation	√	√	√	√
Diacritization	√	√	-	√
POS Tagging	√	√	√	√
Tokenization/ Stemming	√	√	√	√
Lemmatization	√	√	-	√
NER	√	√	√	√
Parsing	√	-	√	√

5 ONTOLOGY FOR THE HADITH

Hadith serves as a guide or explanation of the Quran, so it's crucial to study it carefully to avoid misunderstandings. Ontology, on the other hand, assists in organizing knowledge by categorizing things based on their fundamental characteristics. This hierarchical structuring can be useful for organizing the extensive branches of knowledge within the domain of Hadith.

Based on research studies like [17], [35] and [36], it's crucial to establish an electronic system that makes it easy to access and retrieve Hadiths. This system should also clarify the relationships between different Hadiths and how they relate to each other. However, the current number of works on Islamic Semantic Web applications is insufficient. Therefore, adopting an ontology-based approach to support semantic search from Hadiths can greatly enhance Islamic electronic content and improve information retrieval using the semantic web.

Azmi et al. [2] presented a survey for Computational and natural language processing based studies of hadith. They showed that creating a comprehensive Islamic knowledge ontology is a challenging endeavor, demanding expertise in various Islamic resources.

Bounhas [19] conducted a literature review focusing on computer science research applied to Hadith. The review encompassed various fields, including Natural Language Processing (NLP), Information Retrieval (IR), and Knowledge Extraction (KE). Through this study, existing works were examined and compared. However, a significant challenge arose due to the use of diverse collections of Hadiths in previous studies, making it difficult to objectively compare their results.

Authors in [17] suggest that ontology can aid in storing Hadith commentary, which is currently unavailable. The introduction of a Hadith commentary ontology could enhance ontological-based repositories, facilitating the creation of an online Hadith corpus and establishing connections among Islamic-related ontologies. Figure 3 shows overview of ontological entities and their interconnected relationships.

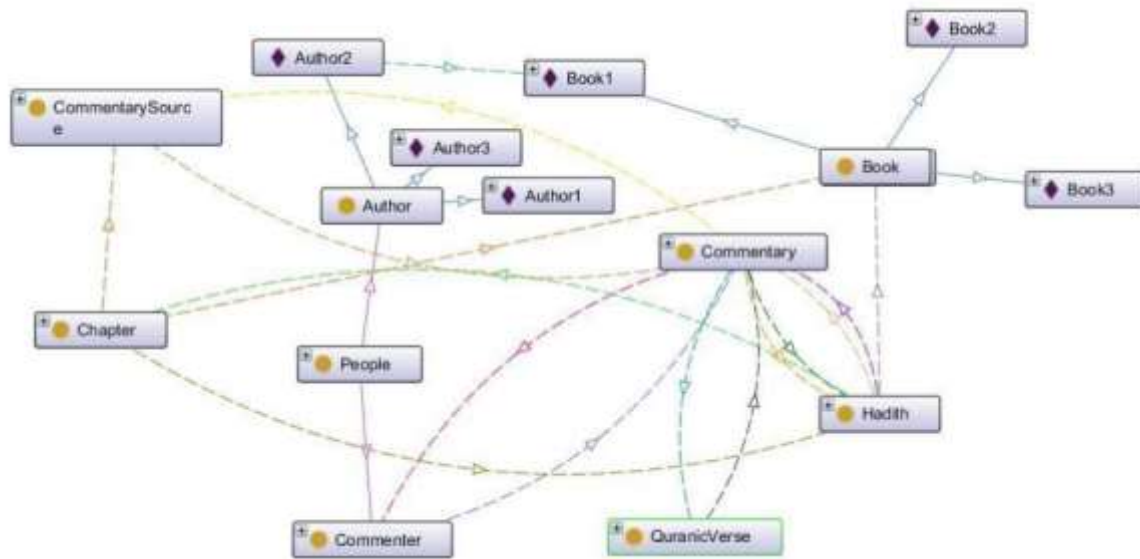


Figure 3 Overview of ontological entities and their interconnected relationships [17]

Furthermore, the authors in [12] propose the development of an ontology-based Isnad Judgment System (IJS) capable of automatically generating suggested judgments for Hadith Isnad based on the rules followed by Hadith scholars. A prototype of this approach has been implemented to demonstrate its feasibility and verify its accuracy.

Additionally, the authors in [12] emphasize that constructing a domain-specific ontology, known as the Hadith Isnad Ontology, could aid in authenticating or judging Isnad. Through evaluation using Hadith examples and DL-Queries, this ontology could potentially generate suggested judgments of Hadith Isnad automatically in future work. Thus, examining Hadith from an ontological perspective offers valuable insights.

The paper [35] highlighted the importance of constructing an ontology to represent the meaning of Hadiths and extract knowledge from the extensive textual sources available in Arabic. This ontology would facilitate visualizing overlaps between Hadith keywords and concepts in the Quran ontology [13]. Additionally, the proposed ontology could potentially be expanded in the future to automatically suggest treatments for specific illnesses based on the Prophet's actions [37]. Figure 4 shows an example of the conceptual model for the hadith ontology as in [35].

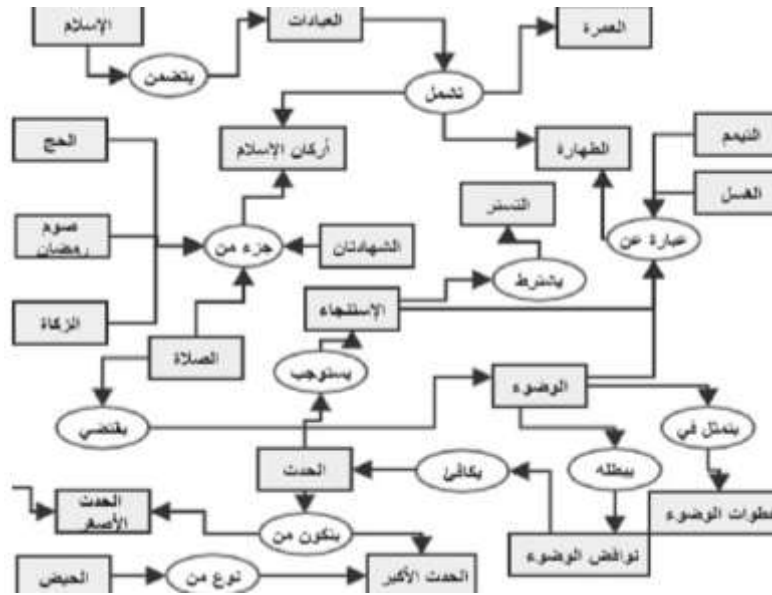


Figure 4 An Example of the Conceptual Model for the Hadith Ontology [35].

Saad et al. in [36] established a framework to extract Islamic concepts, which was later utilized to construct an Islamic knowledge ontology. The researchers introduced an approach for automatically generating instances for the ontology from the Holy Qur'an, employing a combination of Natural Language Processing (NLP), information extraction, and text mining techniques. Their focus was on developing an ontology for Qur'anic verses related to the topic of prayer, known as 'salat.' The extraction process involved identifying verses containing the keyword "salat," followed by verification of these verses with the contents of the Surahs to elaborate on their meanings and establish relevant concepts and relations. Subsequently, the identified pattern was used to extract the verses from the text. In their future work, the authors plan to incorporate hadiths into their framework, as many Qur'anic verses require context from hadiths for proper understanding. Currently, the system's incompleteness affects the presented results, and they may not accurately reflect its actual performance. It is important to note that the entire study relies on the English translation of the Qur'an instead of the original Arabic, which would have posed a greater challenge. The Arabic alphabet lacks case distinctions, yet in this research, approximately 54% of the extracted concepts for the ontology were derived from capitalized letters in the English translation. This highlights the complexity of the task and the need for further refinement when dealing with Arabic text.

Harrag et al. [38] employed association rules to construct an ontology for specific hadiths from Sahih of Bukhari. Ontologies serve as a means to represent knowledge in a manner understandable to both humans and machines, accomplished by encapsulating the semantic aspects of concepts within a particular domain. The researchers utilized association rules, a potent data mining technique, to extract an ontology related to Islamic jurisprudence (fiqh) from the chosen hadiths. Specifically, they employed the Apriori algorithm to develop the proposed ontology. While the study is commendable, it falls short in not conducting experiments to evaluate the efficacy of the created ontology.

Table 4 shows a summary of some works on the ontology for the Hadith. From this survey on the ontology for the Hadith, we concluded that the conceptualization of ontology in the domain of Hadith has received considerable attention, with potential for further advancements in the future to benefit others in this field.

TABLE 4
Summary of Some Works on The Ontology for The Hadith

Ref	Title	Aim	Methodology	Discussion of results
[35]	An Ontological Model of Hadith Text	To allow different kinds of computer systems to use this knowledge effectively, several researchers employed a formal representation of the meaning of Hadith.	An approach influenced by the "METHONTOLOGY" methodology.	Building an ontology that captures the meaning of Hadiths and the valuable knowledge contained within these extensive Arabic textual sources.
[13]	Towards a Joint Ontology of Quran and Hadith	The goal of this study is to list various ontologies and assess them using a corpus-based method to illustrate the connections between these ontologies and the Hadith.	Analysis of documents.	The visualization demonstrates the intersections between the keywords found in the Hadith and the concepts outlined in the Quran ontology.
[17]	Hadith Commentary Repository: An Ontological Approach	The ontology's capability to provide solutions or meet the needs of specific questions.	Assessing based on the methodology explained and published by [39].	An ontology can help store Hadith Commentary, which isn't currently accessible. We predict that once the Hadith commentary ontology becomes available, it will further enhance ontological applications.
[37]	TibbOnto: Knowledge Representation of Prophet Medicine (Tibb Al-Nabawi)	This study's goal is to create a special type of ontology called the Tibb Al-Nabawi ontology. It will organize information about the medicine used by the Prophet in a clear and organized way.	Analyzing documents that are based on genuine Tibb Al-Nabawi Hadith.	In the future, the suggested ontology could expand to automatically create treatments for particular diseases based on the Prophet's practices.
[40]	Towards Building a Prophetic Ontology Based on Quran and Hadith	To test the proposed ontology, we created the Prophetic Question Answering Interface to Ontology (PQuestIO). It's a basic Question Answering (QA) system built in Java.	Constructing the prophetic ontology involves employing ontology development methods and tools like OWL-API, DL-Query, and Protégé.	The system permits users to engage with the ontology to inquire about Prophets and receive answers to their questions from two reliable sources: the Holy Quran and the Hadith of Prophet Mohammad (PBUH) as narrated in Sahih Al-Bukhari, a respected compilation of Hadith.
[41]	An Ontology Based Approach to Enhance Information Retrieval from Al-Shamelah Digital Library	This study aims to use ontology-based annotations to generate search results that are highly relevant and to provide recommendations on related topics.	The structure and layout of OntoADL	Highlighting the role of ontology-based reasoning in delivering intelligent outcomes that align with the user's interests.

(Continued on the next page...)

TABLE 4
Summary of Some Works on The Ontology for The Hadith (Continued)

Ref	Title	Aim	Methodology	Discussion of results
[12]	Building Hadith Ontology to Support the Authenticity of Isnad	The objective of this study is to create a specialized ontology (Hadith Isnad Ontology) to aid in authenticating or evaluating Isnads.	Analysis of documents through a qualitative lens.	Assess the ontology using examples from Hadith and DL-Queries. In future work, the created ontology will be utilized to automatically generate a suggested evaluation of Hadith Isnad.
[14]	An Ontology-Based Approach to Support the Process of Judging Hadith Isnad	This study aims to advance this field through its research efforts. We create an ontology-based Isnad Judgment System (IJS) designed to automatically produce a suggested evaluation of Hadith Isnad.	Both qualitative and quantitative.	Develop an Isnad Judgment System (IJS) based on ontology, which automatically provides a suggested evaluation of Hadith Isnad.
[38]	Ontology extraction approach for prophetic narration (hadith) using association rules	The objective of this study to construct an ontology for specific hadiths from Sahih of Bukhari.	The researchers utilized association rules and employed the Apriori algorithm	While the study is commendable, it falls short in not conducting experiments to evaluate the efficacy of the created ontology.

6 CONCLUSION AND FUTURE WORK

this paper has provided an overview of the importance of Hadith texts in Islamic scholarship and the challenges involved in their analysis and interpretation. It has discussed the complexities of the Arabic language and the development of ontologies tailored to the domain of Hadith. Various research works focusing on computational analysis, ontology development, and application of ontologies in Hadith studies have been reviewed. These works demonstrate the potential of computational methods and ontological approaches in facilitating the study, organization, and retrieval of Hadith information. However, there remain challenges such as ensuring the accuracy and reliability of extracted information, addressing linguistic complexities, and expanding the scope of ontological representations. Future research should focus on refining existing methodologies, integrating additional linguistic and semantic resources, and developing comprehensive ontological frameworks to further advance the field of Hadith studies.

In the future, researchers aim to improve methods for analyzing Hadith texts, ensuring they accurately capture the nuances of the Arabic language. They also plan to expand existing Hadith ontologies to include more comprehensive semantic relationships and contextual information. Collaboration between experts in Islamic studies and computational linguistics will be crucial for advancing these ontological frameworks. Additionally, integrating technologies like machine learning and the semantic web could open up new possibilities for enhancing the accessibility and understanding of Hadith texts, benefiting education, research, and information retrieval in the Islamic scholarly community and beyond.

REFERENCES

- [1] S. H. Altammami, "Artificial intelligence for understanding the Hadith." University of Leeds, 2023.
- [2] A. M. Azmi, A. O. Al-Qabbany, and A. Hussain, "Computational and natural language processing based studies of hadith literature: a survey," *Artif. Intell. Rev.*, vol. 52, pp. 1369–1414, 2019.
- [3] M. M. A. Najeeb, "A novel hadith processing approach based on genetic algorithms," *IEEE Access*, vol. 8, pp. 20233–20244, 2020.
- [4] N. Z. M. Ismail, N. M. Ghazali, and S. F. M. Tawil, "Systematic Literature Review On The Ontology Representation Related The Hadith Corpus," *J. Hadith Stud.*, vol. 8, no. 2, pp. 105:116, 2023.
- [5] F. Harrag, E. El-Qawasmah, and A. M. S. Al-Salman, "Stemming as a feature reduction technique for arabic text categorization," in *2011 10th international symposium on programming and systems*, Algiers, Algeria, IEEE, 2011,

- pp. 128–133.
- [6] M. Naji Al-Kabi, G. Kanaan, R. Al-Shalabi, S. I. Al-Sinjilawi, and R. S. Al-Mustafa, “Al-Hadith text classifier,” *J. Appl. Sci.*, vol. 5, no. 3, pp. 584–587, 2005.
 - [7] “3arabic | Observations in Arabic and Semitic linguistics.” <https://arabicloom.wordpress.com/> (accessed Jan. 30, 2024).
 - [8] S. Saad, N. Salim, H. Zainal, and S. A. M. Noah, “A framework for Islamic knowledge via ontology representation,” in *2010 International Conference on Information Retrieval & Knowledge Management (CAMP)*, Shah Alam, Malaysia, IEEE, 2010, pp. 310–314.
 - [9] A. Kilgarriff and C. Yallop, “What’s in a Thesaurus?,” In *Proceedings of the 2nd Conference on Language Resources and Evaluation (LREC, Athens, Greece)*, 2000, pp. 1371–1379.
 - [10] L. Abouenour, K. Bouzoubaa, and P. Rosso, “Improving Q/A using Arabic wordnet,” in *Proc. The 2008 International Arab Conference on Information Technology (ACIT’2008), Tunisia, December, 2008*.
 - [11] M. Alkhatib, A. A. Monem, and K. Shaalan, “A Rich Arabic WordNet Resource for Al-Hadith Al-Shareef,” *Procedia Comput. Sci.*, vol. 117, pp. 101–110, 2017.
 - [12] R. S. Baraka and Y. Dalloul, “Building Hadith ontology to support the authenticity of Isnad,” *Int. J. Islam. Appl. Comput. Sci. Technol.*, vol. 2, no. 1, pp. 25–39, 2014.
 - [13] S. Altammami, E. Atwell, and A. Alsalka, “Towards a joint ontology of quran and hadith,” *International Journal on Islamic Applications in Computer Science And Technology*, 9, 01–12. 140, 168
 - [14] Y. M. Dalloul, “An ontology-based approach to support the process of judging Hadith Isnad,” *Master’s thesis, Islam. Univ. Gaza*, 2013.
 - [15] A. Shalahuddin and A. Suyadi, “Ulumul Hadits,” *Bandung: Pustaka Setia*, 2008.
 - [16] A. Fageh, “Analisa ontologi dan epistemologi hadis perspektif Ibn Taimiyah,” *TARBIYA Islam. J. Pendidik. Dan Keislam.*, vol. 8, no. 2, pp. 134–148, 2019.
 - [17] A. H. Jaafar and N. Che Pa, “Hadith commentary repository: An ontological approach,” 2016.
 - [18] A. Basid, “Takhrīj Al-H? Adīth Dalam Perspektif Kajian Ontologi Dan Aksiologi Ilmu Hadis,” *Tafāqquh J. Penelit. Dan Kaji. Keislam.*, vol. 4, no. 1, pp. 78–91, 2016.
 - [19] I. Bounhas, “On the usage of a classical Arabic corpus as a language resource: related research and key challenges,” *ACM Trans. Asian low-resource Lang. Inf. Process.*, vol. 18, no. 3, pp. 1–45, 2019.
 - [20] K. Darwish, N. Habash, M. Abbas, H. Al-Khalifa, H. T. Al-Natsheh, H. Bouamor, K. Bouzoubaa, V. Cavalli-Sforza, S. R. El-Beltagy, and W. El-Hajj, “A panoramic survey of natural language processing in the Arab world,” *Commun. ACM*, vol. 64, no. 4, pp. 72–81, 2021.
 - [21] “Farasa.” <https://farasa.qcri.org/> (accessed Mar. 04, 2024).
 - [22] A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak, “Farasa: A fast and furious segmenter for arabic,” in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Demonstrations*, 2016, pp. 11–16.
 - [23] O. Obeid, N. Zalmout, S. Khalifa, D. Taji, M. Oudah, B. Alhafni, G. Inoue, F. Eryani, A. Erdmann, and N. Habash, “CAMEL tools: An open source python toolkit for Arabic natural language processing,” in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 7022–7032.
 - [24] “CAMEL Tools Documentation — camel_tools 1.5.2 documentation.” <https://camel-tools.readthedocs.io/en/latest/> (accessed Mar. 04, 2024).
 - [25] S. Altammami, E. Atwell, and A. Alsalka, “The Arabic-English parallel corpus of authentic hadith,” *Int. J. Islam. Appl. Comput. Sci. Technol.*, vol. 8, no. 2, 2020.
 - [26] M. Mghari, O. Bouras, and A. El Hibaoui, “Sanadset 650K: Data on Hadith narrators,” *Data Br.*, vol. 44, p. 108540, 2022, doi: <https://doi.org/10.1016/j.dib.2022.108540>.
 - [27] “Overview - CoreNLP.” <https://stanfordnlp.github.io/CoreNLP/> (accessed Mar. 04, 2024).
 - [28] “NLTK :: Natural Language Toolkit.” <https://www.nltk.org/> (accessed Mar. 04, 2024).
 - [29] “spaCy · Industrial-strength Natural Language Processing in Python.” <https://spacy.io/> (accessed Mar. 04, 2024).
 - [30] “Overview - Stanza.” <https://stanfordnlp.github.io/stanza/> (accessed Mar. 04, 2024).
 - [31] “Gensim: Topic modelling for humans.” <https://radimrehurek.com/gensim/> (accessed Mar. 04, 2024).
 - [32] “Buckwalter Arabic Morphological Analyzer Version 2.0 - Linguistic Data Consortium.” <https://catalog.ldc.upenn.edu/LDC2004L02> (accessed Mar. 04, 2024).
 - [33] “MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic - ACL Anthology.” <https://aclanthology.org/L14-1479/> (accessed Mar. 04, 2024).
 - [34] A. Pasha, M. Al-Badrashiny, M. T. Diab, A. El Kholy, R. Eskander, N. Habash, M. Pooleery, O. Rambow, and R. Roth, “Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic,” in *Lrec*, 2014, pp. 1094–1101.
 - [35] B. Fairouz, N. Taleb, and A. N. Arari, “An Ontological Model of Hadith Texts,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, 2020.
 - [36] S. Saad, N. Salim, and H. Zainal, “Islamic knowledge ontology creation,” in *2009 International Conference for Internet Technology and Secured Transactions (ICITST)*, London, UK, IEEE, 2009, pp. 1–6.

- [37] A. Al-Rumkhani, M. Al-Razgan, and A. Al-Faris, "TibbOnto: knowledge representation of prophet medicine (Tibb Al-Nabawi)," *Procedia Comput. Sci.*, vol. 82, pp. 138–142, 2016.
- [38] F. Harrag, A. Alothaim, A. Abanmy, F. Alomaigan, and S. Alsalehi, "Ontology extraction approach for prophetic narration (hadith) using association rules," *Int. J. Islam. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 48–57, 2013.
- [39] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology." Stanford knowledge systems laboratory technical report KSL-01-05, Stanford University, Stanford, CA, 94305, March 2001.
- [40] H. A. Al-Sanasleh, "Towards Building a Prophetic Ontology Based on Quran and Hadith." Princess Sumaya University for Technology (Jordan), 2016.
- [41] I. M. AlAgha and M. G. Al-Masri, "An Ontology Based Approach to Enhance Information Retrieval from Al-Shamelah Digital Librar," *IUG J. Nat. Stud.*, vol. 24, no. 1, 2016.

BIOGRAPHY



Mariam Muhammad is an Assistant Teacher at the Department of Computer Sciences, Faculty of Statistical Studies and Research (FSSR), Cairo University. She received her B.Sc. in information systems from faculty of computers and information, Helwan University, Egypt. She received her MCs degree in Information Systems from FSSR, Cairo University, Egypt in 2016.



Shahira Shaaban Azab received the M.Sc. and Ph.D. degrees from Cairo University, Cairo, Egypt, in 2010 and 2018, respectively. She is currently an Assistant Professor with the Computer Science Department, Faculty of Graduate Studies of Statistical Research, Cairo University. Her research interests include machine learning, deep learning, optimization, and natural language processing.



Nesrine Ali Abd el Azim is a lecturer at the Department of Information Systems, (FSSR), Cairo University. She received her B.Sc. in information systems from faculty of computers and information, Cairo University, Egypt. She received her MCs degree in Information Systems from the Faculty of computers and information, Cairo University, Egypt. She received her Ph.D. degree in Information Systems, FSSR, Cairo University in 2015.



Mervat Hassan Gheith is a professor at the Department of Computer Sciences, Faculty of Statistical Studies and Research (FSSR), Cairo University. Dr. Mervat was the head of the department. She published many papers in computer sciences fields such as NLP and other artificial intelligence applications.

ARABIC ABSTRACT

الأنطولوجيا العربية لنصوص الحديث – استعراض للأعمال السابقة

مريم عادل محمد^{1*}، شهيرة شعبان عزب^{2**}، نسرين على عبدالعظيم^{3*}، مرفت حسن غيث^{4**}

*قسم نظم وتكنولوجيا المعلومات، كلية الدراسات العليا للبحوث الإحصائية، جامعة القاهرة، مصر.

¹eng.maryamadel@cu.edu.eg

³Nesrineali79@cu.edu.eg

**قسم علوم الحاسب، كلية الدراسات العليا للبحوث الإحصائية، جامعة القاهرة، مصر

²shahiraazazy@cu.edu.eg

⁴mervat_gheith@yahoo.com

ملخص:

تلعب الأحاديث النبوية دورًا هامًا في العالم الإسلامي، حيث تقدم إرشادًا للمسلمين في فهم وممارسة الإسلام. يستكشف هذا البحث أهمية الأحاديث النبوية، وتصنيفها بناءً على مدى موثوقيتها، وهيكل الحديث النموذجي. كما يناقش استخدام الأدوات البرمجية والأنطولوجيا في تحليل الأحاديث النبوية. كما يفحص البحث اللغة العربية، وخصوصًا تعقيداتها والتحديات التي تواجهها في المهام المتعلقة بمعالجة اللغة الطبيعية. اللغة العربية تُصنّف إلى ثلاثة أنواع رئيسية: اللغة الكلاسيكية واللغة الحديثة واللهجات العربية. وتعاني اللغة العربية من العديد من التحديات التي تؤثر في بناء الأنطولوجيا، على سبيل المثال، نقص الموارد اللغوية وكذلك التشكيل المعقد. كما يستكشف البحث أيضًا تطوير الأنطولوجيات في الدراسات الإسلامية، مسلطًا الضوء على خصائصها الفريدة وتطبيقاتها. وتناقش العديد من الأبحاث المتعلقة بأنطولوجيا الحديث النبوي وتطبيقاتها المحتملة، بما في ذلك إنشاء شبكة الكلمات النبوية، والتحليل الحسابي للأحاديث النبوية، وتطوير أنظمة مبنية على الأنطولوجيا لمصادقة واسترجاع معلومات الأحاديث النبوية.

الكلمات المفتاحية:

الحديث، أنطولوجيا الحديث، أنطولوجيا الدراسات الإسلامية، اللغة العربية