TELKOMNIKA Telecommunication, Computing,Electronics and Control
Vol. 24, No. 3, June 2026, pp. 840∼851
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i3.27447

Computational methodologies for sanad-based hadith analysis: a review

Abdelilah Mhamedi, Mohammed Mghari, Abdelaaziz El Hibaoui
Department of Computer Science, Faculty of Science, Abdelmalek Essaâdi University, Tétouan, Morocco

Article Info

Article history:

Received Aug 2, 2025
Revised Dec 11, 2025
Accepted Jan 30, 2026

Keywords:

Authenticity
Hadith
Machine learning
Natural language processing
Network analysis
Ontology
Sanad

ABSTRACT

Hadith literature, a cornerstone of Islamic tradition, critically depends on the *sanad* (chain of narrators) for authentication, a process traditionally requiring profound scholarly expertise. This paper presents a systematic review of computational methodologies designed to enhance and automate sanad analysis, bridging Islamic studies with advanced artificial intelligence (AI). We categorize progress across four key domains: automated authenticity classification, sophisticated narrator network analysis, textual information extraction (e.g., named entity recognition), and the development of specialized datasets and ontologies. Our findings reveal a significant paradigm shift from rule-based systems to advanced machine learning (ML) and deep learning (DL) techniques. This review synthesizes contributions from over 50 studies, highlighting critical challenges including data scarcity, narrator disambiguation, and cross-linguistic resource limitations. We emphasize the novelty of this cross-domain synthesis and discuss how these intelligent systems can be integrated into digital Islamic archives, low-resource mobile hadith applications, and embedded natural language processing (NLP) engines. This work charts a course for future research to develop more robust, scalable, and ethically grounded computational tools, complementing traditional hadith scholarship with advanced engineering solutions.

Corresponding Author:

Mohammed Mghari
Department of Computer Science, Faculty of Science, Abdelmalek Essaâdi University
P.O. Box. 2121, M'Hannech II, Tétouan, 93030, Morocco
Email: mohammed.mghari@uae.ac.ma

1. INTRODUCTION

Hadith literature, which comprises narrations of the sayings, actions, and approvals of the Prophet Muhammad peace be upon him (PBUH), is a foundational source of Islamic teachings and law, second only to the Qur'an [1], [2]. Every hadith is composed of two primary parts: the *Matn*, which is the text of the narration itself, and the *sanad* (also known as *Isnad*), which is the chain of narrators responsible for transmitting the Matn [3]. The authenticity of a hadith is of utmost importance, as it directly influences Islamic jurisprudence, theology, and daily practice. For centuries, hadith scholars (*muhaddithin*) have employed a rigorous science of hadith criticism (*mustalah al-hadith*) to verify the reliability of these narrations. A central pillar of this science is the meticulous analysis of the sanad [4], [5]. This process involves a deep evaluation of the biographical details of each narrator in the chain (*'ilm al-rijal*) to assess their integrity, memory, and reliability, as well as to ensure the chain is continuous and free from hidden defects [6], [7].

The traditional method of sanad analysis is an exceptionally complex and labor-intensive endeavor, demanding years of specialized study and access to vast biographical and historical resources [8], [9]. The immense volume of hadith literature, which includes hundreds of thousands of narrations across numerous collections, presents a formidable challenge for comprehensive manual analysis. This difficulty is compounded by the intricacies of narrator names, which often appear in multiple variations, and the complex, often branching chains of transmission [10]. In the digital era, while many hadith collections are now accessible online, the computational tools to analyze them are still developing and remain somewhat fragmented. There is a pressing need to harness modern technology to assist scholars, automate repetitive analytical tasks, and uncover new insights from the vast data embedded within the sanads [11], [12].

The convergence of computer science and Islamic studies has catalyzed a vibrant new field dedicated to applying computational techniques to hadith analysis [13]-[15]. This paper offers a systematic review of contemporary sanad-based hadith studies, providing a structured overview of the current state-of-the-art. Recent research demonstrates a clear evolution from early rule-based systems to the adoption of machine learning (ML) and, more recently, advanced deep learning (DL) approaches [8], [16], [17]. By organizing recent contributions into four primary categories automated classification, narrator network analysis, textual component extraction, and dataset/ontology construction this review consolidates current knowledge. We highlight the progression of methodologies, from foundational expert systems to transformer models like bidirectional encoder representations from transformers (BERT) [18]-[20], and synthesize their reported effectiveness and limitations. This review aims to answer the question: ”What are the current computational methodologies applied to sanad-based hadith analysis, what are their reported performances and limitations, and what are the promising future directions?” Our primary focus is on synthesizing the existing tools and theoretical developments that leverage computational power to dissect sanad structures, assess narrator credibility, and enhance scholarly research. We emphasize the novelty of this cross-domain synthesis, showcasing how intelligent systems, can revolutionize hadith studies. Charting a course for future interdisciplinary research will accelerate the development of robust and scalable computational solutions that complement, rather than replace, the invaluable work of traditional hadith scholarship.

2. METHOD

This systematic literature review was conducted to identify, evaluate, and synthesize recent research on computational, sanad-based hadith analysis. The process followed a structured protocol to ensure a comprehensive and unbiased overview of the current state of the field. The search was performed on major academic databases, including IEEE Xplore, ACM Digital Library, Scopus, Google Scholar, and the preprint repository arXiv. The search was conducted using a combination of keywords designed to capture the relevant literature across computer science and Islamic studies.

The primary keywords included: ”hadith sanad analysis”, ”hadith classification”, ”automated hadith authentication”, ”hadith narrator network”, ”social network analysis hadith” [13], [17], ”sanad graph” [15], [21], ”NLP hadith”, ”sanad extraction”, ”narrator name disambiguation”, ”hadith ontology” [16], [22], and ”hadith dataset” [12], [23]-[25]. To ensure the inclusion of the latest advancements, terms such as ”ML”, ”DL”, and ”BERT” [18] were combined with the primary keywords.

Studies were included if they met the following criteria: i) the primary focus was on the sanad (chain of narrators) of hadith; ii) the study applied computational methods, including but not limited to ML, data mining, natural language processing (NLP), or network analysis; iii) the paper was published in English in a peer-reviewed journal, conference proceeding, or as a publicly available technical report or preprint; and iv) the article was published between 2012 and early 2025 to capture a decade of advancements while prioritizing recent work. Papers focusing exclusively on the *Matn* (text) without sanad analysis, purely theological or historical studies without a computational component, and articles not available in full-text were excluded. The initial search yielded numerous articles, which were then screened by title and abstract, followed by a full-text review to determine final eligibility. This process (Figure 1) ensured that the included studies were directly relevant to the scope of this review.

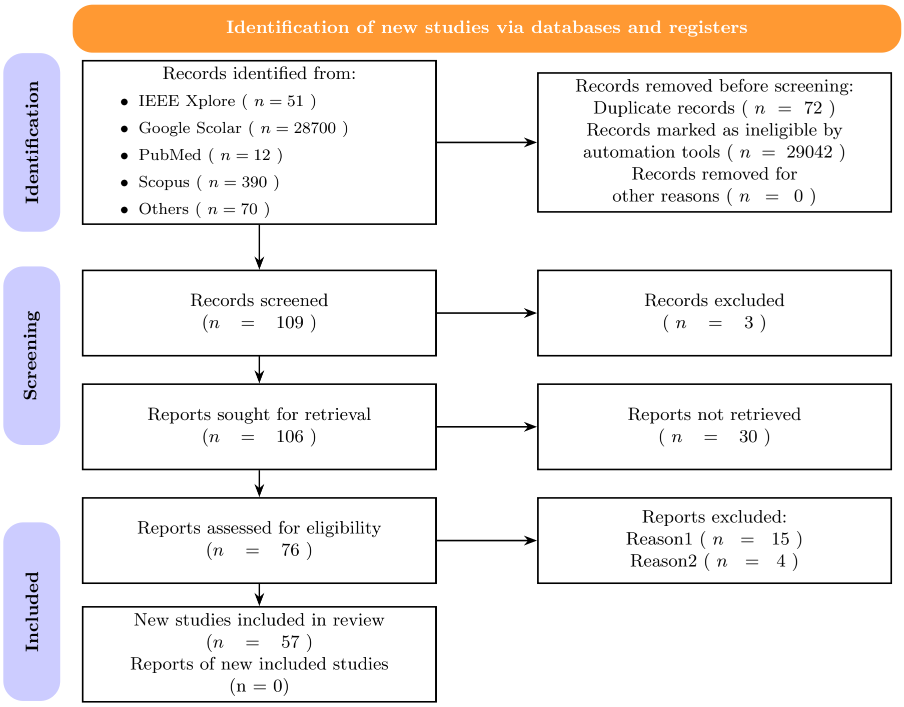

Figure 1. PRISMA low diagram for the systematic review on computational hadith studies

This PRISMA 2020 [26] flow diagram see in Figure 1 systematically maps the process of identifying, screening, and including studies for the systematic literature review on computational tools and techniques in hadith studies. It details the number of records identified from various databases, those removed before screening, and the subsequent stages of record screening, full-text review, and ultimate inclusion in the qualitative synthesis. The scope of the review was rigorously defined by a specific Boolean search query architecture (as illustrated in the query design Figure 2).

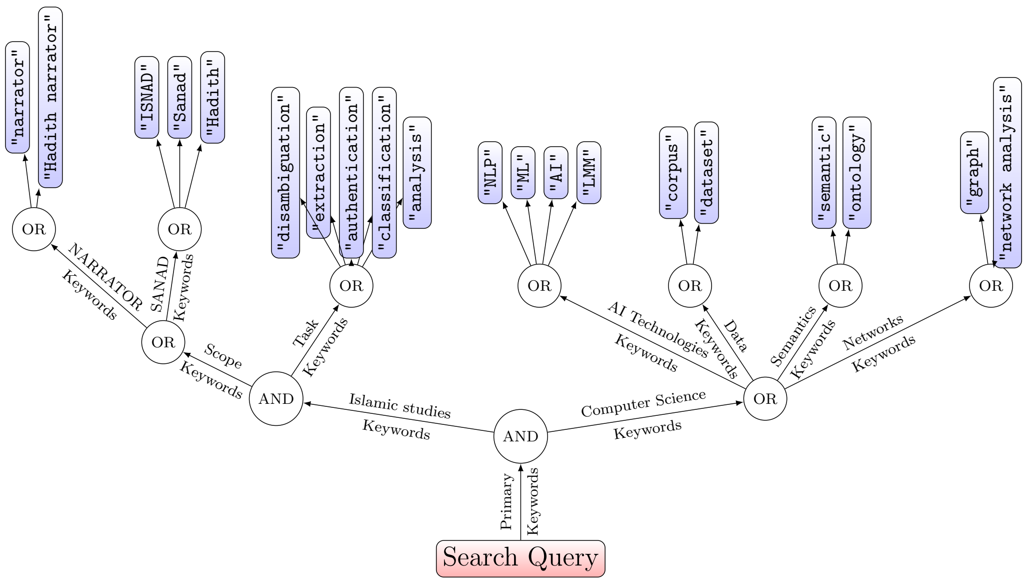

Figure 2. Search query design

This architecture utilized primary keywords connected by the operator AND to ensure that the identified studies operated at the critical intersection of two core domains: Islamic studies and computer science. The architecture, designed using the principle of grouping related terms with OR and enforcing connection using AND, was structured as detailed in the Table 1.

Table 1. Keyword categories and boolean connections

Category (Keywords)	Example terms and purpose	Boolean connection
Sanad/narrator keywords (Islamic studies focus)	Terms like "hadith", "sanad", "hadith narrator", or "narrator" were used to ensure focus on the chain of narrators	Connected by OR within the group
Computational/AI technologies keywords (computer science focus)	Terms such as "LMM", "AI", "ML", and "NLP" captured the required computational methodology	Connected by OR within the group
Task keywords	Terms like "analysis", "classification", "authentication", "extraction", or "disambiguation" targeted specific computational goals	Connected by OR within the group
Structure keywords	Terms like "graph", "networks", "ontology", "semantic", "dataset", and "corpus" captured research related to data structure and formal knowledge representation	Connected by OR within the group

Further modifications and adaptations were implemented to align with the specific requirements of each search engine, thereby enabling the refinement of search results either narrowing or broadening their scope as necessary. In this context, field-specific constraints are applied to ensure that key terms (e.g., sanad) are restricted to semantically relevant metadata fields such as title or abstract, while explicitly excluded from non-relevant fields such as author, thus enhancing precision and reducing noise in retrieval.

3. RESULTS AND DISCUSSION

Our review organizes the findings into four principal domains of sanad-based research. Each domain shows a clear methodological evolution from rule-based systems to sophisticated machine and DL approaches (Figure 3) [27]-[57]. This section provides a detailed discussion of the findings within each category, highlighting key techniques, their performance, and prevailing challenges, supplemented by summary tables for clarity.

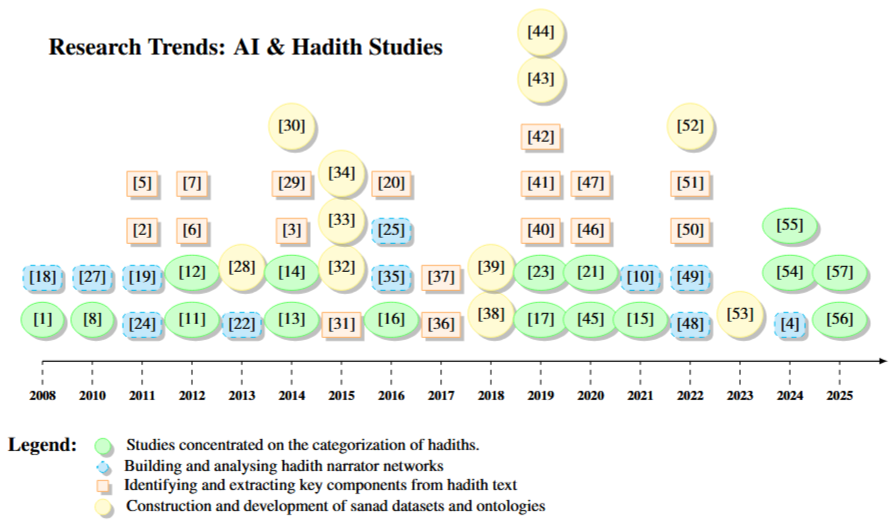

Figure 3. Timeline of sanad-based research, illustrating the methodological evolution in each domain from early rule-based systems to advanced machine and DL approaches

3.1. Automated hadith classification for athenticity assessment

Automating the classification of hadith into traditional categories such as *Sahih* (authentic), *Hasan* (good), *Da'if* (weak), or *Mawdu'* (fabricated) is a primary objective of computational hadith studies [48], [6].

Early attempts in this area relied on fuzzy expert systems and rule-based models that sought to codify the principles of hadith criticism. For example, Ghazizadeh *et al.* [1] proposed a fuzzy expert system to determine hadith validity by modeling parameters like narrator character and sanad continuity, achieving 94% accuracy on a subset of the Shiite collection Al-Kafi. While innovative, these systems were often brittle, collection-specific, and difficult to scale.

The field has since shifted decisively towards ML. Studies have successfully employed supervised learning algorithms where features are derived from the sanad. Aldhlan *et al.* [8], [9], [11] presented a series of papers using decision trees (DT) and Naïve Bayes (NB), incorporating a missing data detector (MDD) to handle incomplete narrator information, which significantly boosted their classification accuracy from 50% to 97% on a dataset of 999 hadiths. Other researchers leveraged heuristic-based systems that assigned weights to narrators based on their rank in classical biographical dictionaries like Ibn Hajar's *Taqrib al-Tahzib*, reporting remarkable accuracies of over 94% on large samples from Sahih Bukhari and Sunan al-Tirmizi [13]. Later work explored vector space models (VSM) and learning vector quantization (LVQ) to consider the order of narrators, achieving 80% precision in distinguishing between Sahih and fabricated hadiths, though performance on Hasan and weak categories was lower [16].

More recently, DL models, particularly those based on the transformer architecture like BERT, are being explored for their ability to process raw text and learn complex representations without manual feature engineering [19], [29]. Abdelaal *et al.* [23] used n-gram techniques (trigrams) and TF-IDF weighting with classifiers like linear SVC, achieving up to 93.69% accuracy. Sentiment analysis has also been applied, where narrator names in the sanad are treated as tokens to predict authenticity, reaching 80% accuracy with a linear SVC model [21]. A 2022 study exploring DL for binary hadith classification (authentic vs. rejected) found that an AraBERT model achieved an accuracy of 91.56% [19]. These results underscore the power of contextual embeddings for capturing the nuanced information within narrator chains. The progression of these methodologies is summarized in Table 2.

Table 2. Studies concentrated on the categorization of hadiths

Ref.	Approach	Preprocessing	Classes	Language	Data source	Metric	Result
[1]	Fuzzy system	-	Valid and not valid	-	El-Kafi	Accuracy	94%
[8]-[11]	DT, NB	-	Valid and not valid	Arabic	999 hadiths	Accuracy	97%
[12]	SaaS, SOA	-	24 classes	-	-	-	-
[13]	HR	Name normalization	Sahih, Hasan, weak	Arabic	Bukhari, Tirmidhi	Accuracy	99%
[14], [15]	ANLP, ANN, SVM, DT, BC	-	Sound, weak	Arabic	-	-	-
[16]	SVM, LVQ	Remove Matn, standardize names	Sahih, Hasan, weak, fabricated	Arabic	160 hadiths	Precision	80%
[17], [23]	DT, NB, linear SVC, SGD, LR	Normalization and tokenization	Sahih, Hasan, weak, fabricated	Arabic	Bukhari and Muslim	Accuracy	up to 93.75%
[45]	Doc2Vec	Stop words and lemmatization	Hadith similarities	Arabic	9 books	Accuracy	80%
[21]	Sentiment analysis	Tokenization, vectorization	Sahih/Hasan, weak	English	Bukhari, Muslim, Tirmidhi	Accuracy	86%
[57]	LR, SVM, RF, AraBERT	REGEX-based preprocessing	Genuine, fake	Arabic	Al-Bukhari, and fake hadiths	F1-score	99.94%
[56]	ArabicBERT, NB, DL, CNN-LSTM	-	Sahih, Hasan, and Da'if and accepted/rejected	English	full-text, sanad-only	F1-score	94.90%
[54]	AraBERT, HistGradientBoostingClassifier	TF-IDF	Three questions with 0,1 or 2 output	Arabic	503 paragraphs	Accuracy	93.16%, 96.55%
[55]	Sanad authentication fuzzy expert system	-	Sahih, Mudallas, Hasan, Matruk, …	Arabic	5,910 chain of narrators	Accuracy	72.2%

3.2. Building and analysing hadith narrator networks

Viewing the entire corpus of hadith transmission as a large-scale social network has opened new avenues for analysis. In this paradigm, narrators are represented as nodes and the transmission of a hadith from one narrator to another is represented as a directed edge. This framework allows for the application of social network analysis (SNA) and graph theory to investigate the structural properties of hadith transmission networks [13], [46].

Early conceptual work proposed modeling narrator chains using graph-theoretic models like directed acyclic graphs (DAGs) to represent the flow of information [18]. Subsequent research focused on creating tools to automatically parse and visualize these chains. Azmi and Badia developed the ”e-Narrator” and ”iTree” systems, which used parsing techniques and domain-specific grammars (EBNF) to generate transmission trees from raw hadith text, achieving an 86.7% success rate on a dataset of 90 hadiths [27], [58], [24]. Other prototypes, like the chain of hadith narrators visualizer (CHN), provided graphical interfaces for students to explore narrator connections [19].

More advanced analyses have applied formal SNA metrics to these narrator graphs. Studies on Sahih Bukhari and Sahih Muslim have used centrality measures (degree, betweenness, PageRank) to identify the most influential narrators in the network, such as Abu Hurayra, Anas bin Malik, and Az-Zuhri, who acted as major hubs in the propagation of knowledge [22], [10], [49]. These findings quantitatively confirm knowledge previously established by classical scholars. Other researchers have used algorithms like SPADE to discover frequent sequential patterns in narrator chains, revealing dominant teacher-student relationships and common transmission pathways [48].

A significant recent trend is the use of graph embedding techniques. Mghari *et al.* [4] introduced Narrator2Vec, a method that learns vector representations (embeddings) of narrators based on their position in the network. These embeddings can be used for tasks such as predicting missing links in a sanad (link prediction), clustering narrators by generation or scholarly affiliation, and identifying narrator similarity. When tested on the all Hadith Corpus, Narrator2Vec achieved 94% accuracy in top-10 narrator prediction tasks. These graph-based approaches provide invaluable macroscopic views of the hadith transmission landscape, highlighting its scale-free nature and identifying key communities and authorities within it. A summary of these analytical approaches is shown in Table 3.

Table 3. Exploring analytical approaches for hadith using sanads

Ref.	Approach	Preprocessing	Classes	Language	Data source	Metric	Result
[18]	DAG	-	Sanad representation	Arabic	-	-	-
[24]-[27]	EBNF	Parse hadith content	Graph, visualization	Arabic	Bukhari	Accuracy	86.7%
[19]	Network graph	-	Graph, visualization	Malay	Nawawi’s 40 hadiths	-	-
[22]	SNA	Extract narrators, remove duplicates	Narrative network analysis	Arabic	Bukhari 5 chapters	Centrality	-
[25], [35]	Network graph	Matn and stop words removal	Graph	Malay	18/30 Hadiths from 9-books/Bukhari	Accuracy	60%
[10]	SNA	-	Narrative network	Arabic	Bukhari	Centrality	-
[48]	SPADE	Transform, clean, format	Sanad analysis	Indonesian	Bukhari	-	-
[49]	SNA	-	Narrative network analysis	English	Muslim	SNA measures	-
[4]	Word embeddings	-	Narrators analysis	Arabic	All hadith corpus	Top-k accuracies	68-94%

3.3. Identifying and extracting key components from hadith text

The automated extraction of key information from hadith text particularly separating the sanad from the Matn and identifying individual narrator names within the sanad is a fundamental task for building structured datasets. This is a challenging NLP problem due to the linguistic characteristics of classical Arabic, the lack of standard punctuation, and the high variability in how narrators' names are cited [50].

Early methods used rule-based approaches and finite state transducers (FST). Harrag *et al.* [2], [3] developed an FST-based entity extractor for Sahih Al-Bukhari, but it struggled with the sanad entity itself, achieving a low F1-score of 33%. Unsupervised tools like the SALAH Project used regular expressions to segment hadith texts, achieving a high effectiveness rate of 97.7% but with limitations on handling complex chains [5]. Other studies employed complex graph transformations and morphological analysis to extract narrator relationships, reporting high precision and recall above 97% for segmentation tasks [6], [7].

The field has seen significant improvement with the adoption of ML-based named entity recognition (NER). Siddiqui *et al.* [29] trained classifiers like Naïve Bayes, DT, and K-nearest neighbors (k-NN) on an annotated corpus to extract narrator names, achieving 90% precision. Najeeb [46], [50] introduced approaches using genetic algorithms (GA) and hidden Markov models (HMM) for sanad processing, reaching accuracies of 81.44% and 86% respectively. N-gram models combined with rule-based filtering have also been used to extract Arabic Person Names, yielding an F-measure of 70.76% [20].

The most significant recent advancements have come from transformer-based models [29]. These models leverage contextual embeddings to achieve state-of-the-art performance. For instance, a study using a semi-supervised BERT model with a feed-forward neural network for NER on Indonesian hadith texts reported an F1-score of 99.27% on a dataset from Bukhari [51]. Another line of research focused on name disambiguation using word sense disambiguation (WSD) techniques combined with a k-NN classifier, reporting an F1-score of 99% on a dataset from Sahih Bukhari, demonstrating a powerful method to resolve narrator ambiguity [59]. These DL techniques are proving indispensable for accurately parsing and structuring hadith texts at scale, a prerequisite for all other higher-level analyses. Table 4 provides an overview of various studies in this category.

Table 4. Studies exploring approaches and techniques for named entity extraction

Ref.	Approach	Preprocessing	Extracted entities	Language	Data source	Metric	Result
[2], [3]	FST-based	-	Sanad extraction	Arabic	Bukhari	F1-score	33%
[5]	RE	-	Sanad	Arabic/English	Bukhari	Accuracy	97.7%
[6], [7]	AMF, FSM, GT	-	Sanad, narrators	Arabic	Ibn Hanbal	F-score	98%
[29]	NER, NB, k-NN, DT	Normalization, stemming	Sanad extraction	Arabic	Bukhari, Musnad Ahmed	Precision	90%
[46]	GA	Sanad/Matn manually divided	Sanad extraction	Arabic	Muslim	Accuracy	81.44%
[50]	HMM, Gazetteer	Sanad/Matn manually divided	Sanad extraction	Arabic	Muslim	Accuracy	86%
[20]	N-gram	-	Narrators extraction	Arabic	6 books	F-measure	65.11%
[31]	POS tags, rule-based	Punctuation removal	Sanad, narrators	Malay	150/1000 from Bukhari	Accuracy	95.3%
[36]	Rule-based, Statistical	Tokenization, Stemming	Narrators	Arabic	Bukhari (prayer)	F1-score	76%
[37]	CRF, FST	-	-	Urdu	Bukhari	F1-score	92.41%
[40]	NER, SVM	Symbol removal, tokenization	Narrators extraction	Indonesian	200 from 9 books	F1-score	90%
[41], [47]	Rule-based, NB	Diacritics/punctuation removal	Sanad extraction	Arabic/English	6 books	Accuracy	92.5%
[42]	HMM	Punctuation removal, Tokenization	Names extraction	Indonesian	9 books	F1-score	86%
[51]	NER, BERT-based	Tokenization	Narrators extraction	Indonesian	102 from Bukhari	F1-score	99.63%

3.4. Construction and development of sanad datasets and ontologies

The foundation for all computational hadith research is the availability of high-quality, structured digital resources. A significant area of work, therefore, involves the construction of comprehensive datasets and formal ontologies to support reproducible research and advanced applications.

Early work focused on creating structured databases from raw text. Bimba *et al.* [34] developed a web-based tool with a relational MySQL database to compile authentic hadith in Malay, linking text to reporter information. Other researchers focused on building lexicons using formalisms like head-driven phrase structure grammar (HPSG) and creating XML-based databases of narrators and their biographical details, often drawing from classical sources like Ibn Hajar's work [32], [33], [60]. The text encoding initiative (TEI) standard was also adopted to normalize and encode hadith texts, with studies developing trigger-word dictionaries to segment Isnad from Matn, achieving a high F-measure of 96% [38], [43], [61].

A major contribution to the field has been the development of large-scale, public datasets. Mahmood *et al.* [39] created a multilingual repository of hadith content extracted from online sources using regular expressions, achieving 100% accuracy for some books. Other projects have focused on creating specialized corpora, such as a non-authentic hadith (NAH) corpus to train models to detect fabricated narrations [44]. Most significantly, recent efforts have produced large-scale datasets for narrator disambiguation, such as the AR-sanad 280 K dataset, which contains 279,625 artificial sanads. Experiments using this dataset with an AraBERT model achieved a 92.9% micro F1 score, demonstrating the value of large-scale synthetic data for training robust models [52].

Alongside datasets, there is growing interest in developing formal ontologies for hadith. Ontologies provide a machine-readable representation of knowledge, defining concepts (e.g., narrator, hadith) and their relationships (e.g., *'narrates'*). These semantic models facilitate advanced querying and logical inference. Dalloul and Baraka created an ontology-based Isnad judgment system that could automatically verify chain continuity based on narrator relationships, achieving 81% accuracy [30], [28]. These resources are foundational for building the next generation of intelligent hadith analysis tools. An overview of these construction efforts is provided in Table 5.

Table 5. Studies on the construction and development of hadith-specific datasets and ontologies

Ref.	Approach	Data type	Language	Data source	Metric	Result
[28], [30]	Sanad ontology	Arabic	6 Books	Accuracy	81%	
[32], [33]	XML, HPSG, Multi-agent	XML, dataset, Isnad tree	Arabic	Bukhari, Muslim	-	-
[34]	MySQL	dataset	Malay	-	-	-
[38], [43]	TEI predefined XML tags	Dataset	Arabic	Bukhari	F-measure	85-96%
[39]	Regular expression	Dataset, XML, CSV	Multilingual	Muslim and Bukhari	F-measure	100%
[44]	-	Dataset, sanad	Arabic	6 books	-	-
[52]	BERT-based (AraBERT)	Dataset, disambiguation	Arabic	6 books	F1-score	92.9%
[53]	RDF, knowledge graph, linked open data	SemanticHadith ontology, knowledge graph	Arabic, Urdu, English	Six hadith collections	-	-

4. CONCLUSION

This comprehensive review has charted the significant progress in the application of computational methods to sanad-based hadith analysis. The field is rapidly maturing, moving from foundational rule-based systems to sophisticated DL and network science methodologies. The analysis of automated classification techniques shows a clear performance advantage for ML over static rules, with recent transformer-based models like BERT setting new benchmarks for authenticity assessment. The exploration of narrator networks through SNA has provided quantitative validation of classical hadith scholarship and offers powerful tools for visualizing and understanding the macro-structure of knowledge transmission in Islam. Concurrently, advancements in NLP have been critical in automating the foundational tasks of sanad segmentation and narrator entity recognition, making large-scale analysis feasible. Finally, the development of large, publicly available datasets and formal ontologies is providing the essential infrastructure to fuel further research and ensure reproducibility. These advancements collectively move the body of scientific knowledge forward by demonstrating the immense potential of interdisciplinary collaboration between computer science and Islamic studies.

Despite this progress, challenges remain. Many developed datasets are still limited in scope, and the models trained on them may not generalize well across different hadith collections. The problem of narrator name disambiguation remains a significant hurdle, though recent graph-based and BERT-powered approaches show great promise. The heavy reliance on regular expressions in some data extraction tasks can be brittle, and the creation of robust, adaptable NLP pipelines is an ongoing area of research. Looking forward, future work should focus on creating larger, more diverse, and standardized benchmark datasets that cover a wider range of hadith literature, including less-canonical works. There is great promise in exploring more advanced graph neural network (GNN) architectures for narrator network analysis and leveraging multimodal models that can analyze both sanad and Matn concurrently to create a more holistic authenticity assessment. The continued partnership between computer scientists and traditional hadith scholars is essential to ensure that these technological advancements are developed responsibly, rigorously, and in a way that genuinely supports and enhances our understanding of the rich heritage of Islamic traditions.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abdelilah Mhamedi	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓		
Mohammed Mghari	✓	✓			✓	✓				✓	✓			
Abdelaaziz El Hibaoui	✓			✓		✓			✓	✓	✓		✓	

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] M. Ghazizadeh, M. H. Zahedi, M. Kahani, and B. Minaei Bidgoli, "Fuzzy expert system in determining hadith validity," in *Advances in Computer and Information Sciences and Engineering,* T. Sobh, Ed., Springer Netherlands, 2008, pp. 354–359, doi: 10.1007/978-1-4020-8741-7_64.

[2] F. Harrag, E. El-Qawasmeh, and A. M. Salman Al-Salman, "Extracting named entities from prophetic narration texts (Hadith)," in *Communications in Computer and Information Science*, vol. 180 CCIS, no. PART 2, J. M. Zain, W. M. B. Wan Mohd, and E. El-Qawasmeh, Eds., Springer Berlin Heidelberg, 2011, pp. 289–297, doi: 10.1007/978-3-642-22191-0_26.

[3] F. Harrag, "Text mining approach for knowledge extraction in Sahîh Al-Bukhari," *Computers in Human Behavior*, vol. 30, pp. 558–566, 2014, doi: 10.1016/j.chb.2013.06.035.

[4] M. Mghari, O. Bouras, and A. El Hibaoui, "Narrator2Vec: An Efficient Narrator Representation in Hadith Literature Using Word Embedding," *Arabian Journal for Science and Engineering*, vol. 49, no. 3, pp. 4479–4494, 2024, doi: 10.1007/s13369-023-08224-7.

[5] M. Boella, F. R. Romani, A. Al-Raies, C. Solimando, and G. Lancioni, "The SALAH project: Segmentation and linguistic analysis of hadīt Arabic texts," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7097 LNCS, M. V. M. Salem, K. Shaalan, F. Oroumchian, A. Shakery, and H. Khelalfa, Eds., Springer Berlin Heidelberg, 2011, pp. 538–549, doi: 10.1007/978-3-642-25631-8_49.

[6] J. Makhlouta, F. Zaraket, and H. Harkous, "Arabic entity graph extraction using morphology, finite state machines, and graph transformations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7181 LNCS, no. PART 1, A. Gelbukh, Ed., Springer Berlin Heidelberg, 2012, pp. 297–310, doi: 10.1007/978-3-642-28604-9_25.

[7] F. Zaraket and J. Makhlouta, “Arabic cross-document NLP for the hadith and biography literature,” *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference*, FLAIRS-25, pp. 256–261, 2012.

[8] K. A. Aldhlan, A. M. Zeki, and A. M. Zeki, “Datamining and Islamic knowledge extraction: Alhadith as a knowledge resource,” in *Proceeding of the 3rd International Conference on Information and Communication Technology for the Moslem World: ICT Connecting Cultures, ICT4M 2010, IEEE*, 2010, p. H-21-H-25, doi: 10.1109/ICT4M.2010.5971934.

[9] K. Aldhaln, A. Zeki, A. Zeki, and H. Alreshidi, “Improving knowledge extraction of Hadith classifier using decision tree algorithm,” in *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management*, CAMP’12, IEEE, 2012, pp. 148–152, doi: 10.1109/InfRKM.2012.6205024.

[10] T. Alam and J. Schneider, “Social Network Analysis of Hadith Narrators from Sahih Bukhari,” *Proceedings of 2020 7th IEEE International Conference on Behavioural and Social Computing*, BESC 2020, vol. abs/2102.02009, 2020, doi: 10.1109/BESC51023.2020.9348299.

[11] K. A. Aldhlan, A. M. Zeki, A. M. Zeki, and H. A. Alreshidi, “Novel mechanism to improve hadith classifier performance,” in *Proceedings - 2012 International Conference on Advanced Computer Science Applications and Technologies*, ACSAT 2012, IEEE, 2013, pp. 512–517, doi: 10.1109/ACSAT.2012.93.

[12] K. Bilal and S. Mohsin, “Muhadith: A cloud based distributed expert system for classification of Ahadith,” in *Proceedings - 10th International Conference on Frontiers of Information Technology*, FIT 2012, IEEE, 2012, pp. 73–78, doi: 10.1109/FIT.2012.22.

[13] A. M. Azmi, A. M. Alofaidly, ”A novel method to automatically pass hukm on Hadith,” *5th International Conference on Arabic Language Processing (CITALA ’14)*, 2014.

[14] M. M. Najeeb, “Towards Innovative System for Hadith Isnad Processing,” *International Journal of Computer Trends and Technology*, vol. 18, no. 6, pp. 257–259, 2014, doi: 10.14445/22312803/ijctt-v18p154.

[15] M. M. A. Najeeb, “Towards a Deep Leaning-based Approach for Hadith Classification,” *European Journal of Engineering and Technology Research*, vol. 6, no. 3, pp. 9–15, 2021, doi: 10.24018/ejeng.2021.6.3.2378.

[16] M. Ghanem, A. Mouloudi, and M. Mourchid, “Classification of Hadiths using LVQ based on VSM Considering Words Order,” *International Journal of Computer Applications*, vol. 148, no. 4, pp. 25–28, 2016, doi: 10.5120/ijca2016911077.

[17] H. M. Abdelaal and H. A. Youness, “Hadith Classification using Machine Learning Techniques According to its Reliability,” *Romanian Journal of Information Science and Technology*, vol. 22, no. 3–4, pp. 259–271, 2019.

[18] U. Relational and S. I. Hyder, “Towards a Database Oriented Hadith Research Using Relational, Algorithmic and Data-Warehousing Techniques,” *The Islamic Culture, Quarterly Journal of Shaikh Zayed Islamic Center for Islamic and Arabic Studies*, vol. 19, no. March, p. 14, 2008.

[19] Z. Shukur, N. Fabil, J. Salim, and S. A. Noah, “Visualization of the hadith chain of narrators,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, H. B. Zaman, P. Robinson, M. Petrou, P. Olivier, T. K. Shih, S. Velastin, and I. Nyström, Eds., Springer, 2011, pp. 340–347, doi: 10.1007/978-3-642-25200-6_32.

[20] M. Alhawarat, “A domain-based approach to extract Arabic person names using N-grams and simple rules,” *Asian Journal of Information Technology*, vol. 14, no. 8, pp. 287–293, 2015, doi: 10.3923/ajit.2015.287.293.

[21] F. Haque, A. H. Orthy, and S. Siddique, “Hadith Authenticity Prediction using Sentiment Analysis and Machine Learning,” in *14th IEEE International Conference on Application of Information and Communication Technologies, AICT 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc.*, 2020, pp. 1–6, doi: 10.1109/AICT50176.2020.9368569.

[22] M. A. Ahmad, “Towards the Analysis of Narrative Networks,” 2013. Accessed: Oct. 25, 2025. [Online]. Available: http://www.aurumahmad.com/assets/pdf/AhmadNarrative.pdf

[23] H. M. Abdelaal, A. M. Ahmed, W. Ghribi, and H. A. Y. Alansary, “Knowledge Discovery in the Hadith According to the Reliability and Memory of the Reporters Using Machine Learning Techniques,” *IEEE Access*, vol. 7, pp. 157741–157755, 2019, doi: 10.1109/ACCESS.2019.2944118.

[24] A. Azmi and N. AlBadia, “Mining and visualizing the narration tree of hadiths (Prophetic traditions),” *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches*, pp. 239–257, 2011, doi: 10.4018/978-1-61350-447-5.ch016.

[25] N. Alias, N. A. Rahman, N. K. Ismail, Z. M. Nor, and M. N. Alias, Searching Algorithm of Authentic Chain of Narrators’ in *Shahih Bukhari Book*, no. May. 2016.

[26] M. J. Page *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, 2021, doi: 10.1136/bmj.n71.

[27] A. M. Azmi and N. Bin Badia, “e-Narrator - an application for creating an ontology of Hadiths narration tree semantically and graphically,” *Arabian Journal for Science and Engineering*, vol. 35, no. 2 C, pp. 51–68, 2010.

[28] Y. M. Dalloul, “An Ontology-Based Approach to Support the Process of Judging Hadith Isnad,” in *2012 International Conference on Advanced Computer Science Applications and Technologies*, 2013, pp. 1–108.

[29] M. A. Siddiqui, M. E. Saleh, and A. A. Bagais, “Extraction and Visualization of the Chain of Narrators from Hadiths using Named Entity Recognition and Classification,” *International Journal of Computational Linguistics Research*, vol. 5, no. 1, pp. 14–25, 2014.

[30] R. S. Baraka and Y. M. Dalloul, “Building Hadith Ontology to Support the Authenticity of Isnad,” *International Journal on Islamic Applications in Computer Science And Technology*, vol. 2, no. 1, pp. 25–39, 2014.

[31] N. Abd Rahman, N. Alias, N. K. Ismail, Z. Bin M. Nor, and M. N. B. Alias, “An identification of authentic narrator’s name features in Malay hadith texts,” in *ICOS 2015 - 2015 IEEE Conference on Open Systems*, IEEE, 2016, pp. 79–84, doi: 10.1109/ICOS.2015.7377282.

[32] M. Najeeb, A. Abdelkader, M. Al-Zghoul, and A. Osman, “A Lexicon for Hadith Science Based on a Corpus,” *International Journal of Computer Science and Information Technologies*, vol. 6, no. 2, pp. 1336–1340, 2015.

[33] M. M. Najeeb, “Multi-agent system for hadith processing,” *International Journal of Software Engineering and its Applications*, vol. 9, no. 9, pp. 153–166, 2015, doi: 10.14257/ijseia.2015.9.9.13.

[34] A. Bimba, M. A. Ismail, N. Idris, S. Jaafar, and R. Mahmud, ”Towards Enhancing the Compilation of Al-Hadith Text in Malay,” *International Proceedings of Economics Development and Research* vol. 83, no. March. 2015.

[35] N. Alias, N. A. Rahman, N. K. Ismail, Z. M. Nor, and M. N. Alias, “Graph-based text representation for Malay translated hadith text,” in *2016 3rd International Conference on Information Retrieval and Knowledge Management, CAMP 2016 - Conference Proceedings, IEEE*, 2017, pp. 60–66, doi: 10.1109/INFRKM.2016.7806336.

[36] S. S. Balgasem and L. Q. Zakaria, “A hybrid method of rule-based approach and statistical measures for recognizing narrators name in hadith,” in *Proceedings of the 2017 6th International Conference on Electrical Engineering and Informatics: Sustainable Society Through Digital Innovation, ICEEI 2017, IEEE*, 2017, pp. 1–5, doi: 10.1109/ICEEI.2017.8312417.

[37] A. Mahmood, H. U. Khan, Zahoor-Ur-Rehman, and W. Khan, “Query based information retrieval and knowledge extraction using Hadith datasets,” in *Proceedings - 2017 13th International Conference on Emerging Technologies, ICET2017, IEEE*, 2017, pp. 1–6, doi: 10.1109/ICET.2017.8281714.

[38] H. Maraoui, K. Haddar, and L. Romary, “Encoding Prototype of Al-Hadith Al-Shareef in TEI,” in *Communications in Computer and Information Science*, A. Lachkar, K. Bouzoubaa, A. Mazroui, A. Hamdani, and A. Lekhouaja, Eds., Springer International Publishing, 2018, pp. 217–229, doi: 10.1007/978-3-319-73500-9_16.

[39] A. Mahmood, H. U. Khan, F. K. Alarfaj, M. Ramzan, and M. Ilyas, “A multilingual datasets repository of the Hadith content,” International *Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, pp. 165–172, 2018, doi: 10.14569/IJACSA.2018.090224.

[40] F. A. Yusup, M. A. Bijaksana, and A. F. Huda, “Narrator’s name recognition with support vector machine for indexing Indonesian hadith translations,” *Procedia Computer Science*, vol. 157, pp. 191–198, 2019, doi: 10.1016/j.procs.2019.08.157.

[41] S. Altammami, E. Atwell, and A. Alsalka, “Text segmentation using n-grams to annotate Hadith corpus,” in *Proceedings of the 3rd Workshop on Arabic Corpus Linguistics*, M. El-Haj, P. Rayson, E. Atwell, and L. Alsudias, Eds., Association for Computational Linguistics, 2019, pp. 31–39. Accessed: Oct. 25, 2025. [Online]. Available: https://www.aclweb.org/anthology/W19-5605

[42] W. P. Sari, M. A. Bijaksana, and A. F. Huda, “Indexing name in hadith translation using hidden Markov model (HMM),” in *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*, 2019, pp. 1–5, doi: 10.1109/ICoICT.2019.8835296.

[43] H. Maraoui, K. Haddar, and L. Romary, “Segmentation Tool for Hadith Corpus to Generate TEI Encoding,” in *Advances in Intelligent Systems and Computing*, A. E. Hassanien, M. F. Tolba, K. Shaalan, and A. T. Azar, Eds., Springer International Publishing, 2019, pp. 252–260, doi: 10.1007/978-3-319-99010-1_23.

[44] T. Tarmom, E. Atwell, and M. Alsalka, Non-Authentic Hadith Corpus: Design and Methodology, *International Journal on Islamic Applications in Computer Science And Technology*, vol. 8, no. 3, 2020.

[45] A. M. Abdelghany, H. M. Abdelaal, A. M. Kamr, and P. M. Elkafrawy, “Doc2Vec: An approach to identify Hadith Similarities,” *Australian Journal of Basic and Applied Sciences*, no. December 2020, pp. 46–53, 2020.

[46] M. M. A. Najeeb, “A novel hadith processing approach based on genetic algorithms,” *IEEE Access*, vol. 8, pp. 20233–20244, 2020, doi: 10.1109/ACCESS.2020.2968417.

[47] S. Altammami, E. Atwell, and A. Alsalka, “Constructing a bilingual hadith corpus using a segmentation tool,” in *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds., European Language Resources Association, 2020, pp. 3390–3398. [Online]. Available: https://aclanthology.org/2020.lrec-1.415/

[48] R. Yotenka, S. K. Dini, A. Fauzan, and A. Ahdika, “Exploring the relationship between hadith narrators in Book of Bukhari through SPADE algorithm,” *MethodsX*, vol. 9, p. 101850, 2022, doi: 10.1016/j.mex.2022.101850.

[49] S. Saeed, S. Yousuf, F. Khan, and Q. Rajput, “Social network analysis of Hadith narrators,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3766–3774, 2022, doi: 10.1016/j.jksuci.2021.01.019.

[50] M. M. A. Najeeb, “A Hidden Markov Model-Based Tagging Approach for Arabic Isnads of Hadiths,” Mathematical Problems in Engineering, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/7160509.

[51] E. T. Luthfi, Z. I. M. Yusoh, and B. M. Aboobaider, “BERT based Named Entity Recognition for Automated Hadith Narrator Identification,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 604–611, 2022, doi: 10.14569/IJACSA.2022.0130173.

[52] S. Mahmoud, O. Saif, E. Nabil, M. Abdeen, M. Elnainay, and M. Torki, “AR-Sanad 280K: A Novel 280K Artificial Sanads Dataset for Hadith Narrator Disambiguation,” *Information (Switzerland)*, vol. 13, no. 2, 2022, doi: 10.3390/info13020055.

[53] A. B. Kamran, B. Abro, and A. Basharat, “SemanticHadith: An ontology-driven knowledge graph for the hadith corpus,” *Journal of Web Semantics*, vol. 78, p. 100797, 2023, doi: 10.1016/j.websem.2023.100797.

[54] M. Hendawi, M. Torki, A. Elmasry, and A. Khalafallah, “Automating Sanad Continuity Verification in Disconnected Hadith Using Machine Learning,” in *34th International Conference on Computer Theory and Applications, ICCTA 2024, Institute of Electrical and Electronics Engineers Inc.*, 2024, pp. 206–212, doi: 10.1109/ICCTA64612.2024.10974901.

[55] E. T. Luthfi and Z. I. M. Yusoh, “the Fuzzy Rule Based System for Determining the Level of Hadith Sanad Authentication Status,” *International Journal of Innovative Computing, Information and Control*, vol. 20, no. 5, pp. 1463–1477, 2024, doi: 10.24507/ijicic.20.05.1463.

[56] K. R. Firmansyah and K. M. Lhaksmana, “Classification of Hadith Authenticity Using ArabicBERT,” in *2025 International Conference on Data Science and Its Applications, ICoDSA 2025, IEEE*, 2025, pp. 1003–1008, doi: 10.1109/ICoDSA67155.2025.11157495.

[57] J. Alghamdi, A. Albukhari, and T. Al-Dala’in, “Pretrained Models Against Traditional Machine Learning for Detecting Fake Hadith,” *Electronics (Switzerland)*, vol. 14, no. 17, 2025, doi: 10.3390/electronics14173484.

[58] A. Azmi and N. Bin Badia, “iTree - Automating the construction of the narration tree of Hadiths (prophetic traditions),” in *Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 2010, IEEE*, 2010, pp. 1–7, doi: 10.1109/NLPKE.2010.5587810.

[59] A. Prasetio, M. Bijaksana, and A. Suryani, “Name Disambiguation Analysis Using the Word Sense Disambiguation Method in Hadith,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 4, no. 2, pp. 68–74, 2020, doi: 10.29408/edumatic.v4i2.2551.

[60] M. M. A. Najeeb, “XML database for Hadith and narrators,” *American Journal of Applied Sciences*, vol. 13, no. 1, pp. 55–63, 2016, doi: 10.3844/ajassp.2016.55.63.

[61] H. Maraoui, K. Haddar, and L. Romary, “Modeling of Al-Hadith Al-Shareef with TEI,” in *Proceedings - 2017 International Conference on Engineering and MIS, ICEMIS 2017*, 2017, pp. 1–5, doi: 10.1109/ICEMIS.2017.8273053.

BIOGRAPHIES OF AUTHORS

Abdelilah Mhamedi is a Professor of Computer Science in the Classes Préparatoires aux Grandes Écoles (CPGE) since September 2014, with a primary focus on delivering comprehensive instruction in computer science. During this tenure, he actively contributed to the development and delivery of professional training for educators and assumed coordination responsibilities for working teams. Academically, he holds a Master's degree in Information Systems Engineering from the Faculty of Sciences Semlalia, Cadi Ayyad University Marrakech. His current research endeavors are concentrated on artificial intelligence (AI), machine learning applications, large language model (LLM) applications, smart cities, and smart grids. He can be contacted at email: abdelilah.mhamedi@etu.uae.ac.ma.

Mohammed Mghari received the Ph.D. degree in AI and Computer Science from the Abdelmalek Essaâdi University in 2024. His Ph.D. work specifically involved applying advanced transformer models to low-resource languages, achieving state-of-the-art performance on a specific task. He is currently a postdoc researcher at the Faculty of Sciences at Abdelmalek Essaâdi University, Tetouan. His current research interests are focused on the intersection of deep learning and computer vision, while also maintaining an interest in artificial intelligence, machine learning, and natural language processing. He can be contacted at email: mohammed.mghari@uae.ac.ma.

Abdelaaziz El Hibaoui is a Full Professor of Computer Sciences at Abdelmalek Essaadi University, Morocco, where he has also headed the Computer Sciences and Systems Engineering Laboratory since 2017. He earned his Ph.D. from Bordeaux 1 University, France, in 2016. His research focuses on randomized distributed algorithms, privacy and computer security, smart grid, AI/machine learning, and informatics for renewable energy, with numerous contributions to refereed journals and conference proceedings. He has a distinguished record of service, coordinating the Mathematics and Computer Sciences Stream at Abdelmalek Essaadi University and serving in various leadership roles for national and international conferences. He can be contacted at email: aelhibaoui@uae.ac.ma.