Research Article

A Hidden Markov Model-Based Tagging Approach for Arabic Isnads of Hadiths

Moath Mustafa Ahmad Najeeb

Computer Science Department, College of Computing at Al-Qunfudah, Umm Al-Qura University, Mecca 21955, Saudi Arabia

Correspondence should be addressed to Moath Mustafa Ahmad Najeeb; mmnajeeb@uqu.edu.sa

Received 12 November 2021; Accepted 15 January 2022; Published 28 February 2022

Academic Editor: Nianyin Zeng

Copyright © 2022 Moath Mustafa Ahmad Najeeb. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hadith judgment implies checking the validity of Hadith to decide whether it is correct (trustworthy) or false (bogus). “Matn” and “Isnad” are the main constituents of Hadith; “Matn” is the sayings of the prophet, whereas “Isnad” represents the narrators’ series. The first step of Hadith judgment is the extraction of narrators’ names, after that, the rules of judgment, which were set out by Hadith’s scientists, could be implemented, three of these rules are particularly related to the narrators’ series, and these rules are continuity of the transmission chain, the trustworthiness of the narrators, and the preciseness of the narrators. Therefore, to check the authenticity of Hadiths, the three conditions must be satisfied, and to do so, the narrators’ names must be extracted first. Isnad contains many words and phrases called “Isnad-Phrases”; these phrases have many types or categories called part of Isnads (POIs) like Narrator-Name, Prophet-Name, and Received-Method. A lot of computational research studies suggest serving Hadith sciences by extracting the narrators’ names and other POIs using various approaches. This study presents a new hybrid approach founded on the hidden Markov model (HMM) and gazetteer lists to process “Isnad.” The approach objective is to expect all POIs in the Isnad including narrators’ names. The experiments carried on 1,000 Hadiths from “Sahih Muslim”: 900 Hadiths as training dataset and 100 Hadiths as testing dataset, and the results show a noteworthy accuracy for the proposed hybrid approach.

1. Introduction

Hadith is the second enactment root in Islam behind Quran; it represents the entire life of Prophet Muhammad, such as his deeds, sayings, and actions. The main constituents of Hadith are as follows: “Isnad” and “Matn,” and Isnad represents the narrators’ series, whereas “Matn” is the saying of the prophet. Since the first century AH (seventh century CE), Hadith scientists set the rules of many Hadith sciences such as “Al-Jarh Wa Al-Ta’dil” and “Mustalah Al-Hadith.” Hundreds of glossaries and books were written to be in the service of these sciences, and the following are the essential books that gather the correct (authentic) Hadiths: “Sahih Muslim,” “Sahih Al-Bukhari,” “Sunan Ibn Majah,” “Sunan Abu Dawood,” “Sunan An-Nasa’i,” and “Sunan Al-Tirmizi” [1].

Fundamental science in Hadith is “Al-Jarh Wa Al-Ta’dil,” it provides a structural procedure to inspect the biography of Hadith’s narrators, and it involves two

principal sections: “Al-Jarh” (criticism) and “Al-Ta’dil” (praising). Criticism examines the sincerity, trustiness, and honesty of narrators, because they may have a bad memory, fake narrators, unknown narrators, liar narrators, etc. This section ranks narrators into six levels, rank 1 shows that the narrator has less criticism level, the expressions used to indicate this rank are many such as his Hadith is soft, not safe, etc., whereas rank 6 shows that the narrator is completely liar, and the expressions used to indicate this rank are many such as compulsive liar, he is fabricator, he lies, etc.

Section two, praising, examines the trustworthiness of narrators, it ranks narrators into six levels, rank 1 shows that the narrator has the uppermost standard of praise, the expressions used to indicate this rank are many such as most reliable of the people, most established of the people, etc., whereas the rank 6 shows that the narrator has the lowest standard of praise, and the expressions used to indicate this rank are many such as acceptable, satisfactory in Hadith, etc.

TABLE 1: POIs for Hadith1.

Isnad-Phrase	Part of Isnad (POI)
حَدَّثَنَا	Received-Method
قُتَيْبَةُ بْنُ سَعِيدٍ	Narrator-Name
هَدِيدُ بْنُ حَرْبٍ	Narrator-Name
قَالَ	Received-Method-Prefix
حَدَّثَنَا	Received-Method
جَرِيدٌ	Narrator-Name
عَنِ	Received-Method
عُمَارَةَ بْنِ الْقُعُقَاعِ	Narrator-Name
عَنِ	Received-Method
أَبِي	Narrator-Name-Prefix
عَزْرٍ	Narrator-Name
عَنْ	Received-Method
أَبِي	Narrator-Name-Prefix
هُرَيْرَةَ	Narrator-Name
قَالَ	Others
وَلِلَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ	Prophet-Name

TABLE 2: HMM elements.

Symbol	Description
	Number of states
N:	There is a finite set of states in a model. The states in an HMM are hidden, but there is a lot of significance to these states in defining an HMM. The individual states are represented as $S_1, S_2, S_3, \dots, S_N$ $S = \{S_1, S_2, S_3, \dots, S_n\}$.
S:	Space of states $\{S_1, S_2, \dots, S_N\}$ N is the number of states
	Number of observations
M:	It is the number of distinct symbols observable in states. These symbols correspond to the observable output of the system that is being modeled. The individual states are represented as $O_1, O_2, \dots, O_M$
O:	Space of observations $\{O_1, O_2, \dots, O_M\}$
A:	Transmission probability matrix A is the transition array that stores the state transition probabilities $A = \{a_{ij}\}$, where a_{ij} stores the probability of state j following state i
B:	Emission probability matrix After each transition is made, a symbol is an output based on the emission probability matrix, which depends on the current state
π :	Start probability It is the probability of state S_i being the start state in an observation sequence

related works, the proposed approach is explained in Section 3, the experimental results are discussed in Section 4, and finally, the conclusion is presented in Section 5.

2. Related Works

Many computational types of research studies are introduced in the literature to support Arabic natural language processing (ANLP) and Islamic sciences. In [10], Harrag and Hamdi-Cherif presented a classification technique for Hadiths depending on the query of the user. Vector space model, cosine, and TF-IDF are used in this study, it used 60 Hadiths to perform the experiments, and the precision and recall were 66% and 80%, respectively. In [11], the authors introduced a classification system of Hadiths' authenticity depending on the fuzzy technique, the experiments were conducted on Al-Kafi book, the authors did not mention the

number of Hadiths they used in the study, and the accuracy of the system was 94%. In [12], the researchers introduced a system to classify Hadiths depending on the artificial neural network (ANN) and singular value decomposition (SVD), the experiments were conducted on 453 Hadiths, and the F-measures were 85.75% and 88.33% for ANN and ANN with SVD, respectively.

Bounhas et al. [13] proposed a Naïve Bayes (NB) classifier for the narration chain reliability of Hadiths, the experiments were conducted on 1000 Hadiths, and the F-measure was 89.01%. In [14–16], the same researchers introduced a graph construction for narration series depending on context-free grammar (CFG), semantic web ontology, and memory-based learning, the experiments were conducted on 90 Hadiths from Sahih Muslim and Sahih Al-Bukhari, and the success rate was 86.7%. In [17], the authors proposed a new decision tree-based approach to classify

Hadiths; this approach categorizes Hadiths into four collections: “Maudo,” “Daief,” “Hasan,” and “Sahih,” the experiments were conducted on 999 Hadiths from Sahih Al-Bukhari, Sunan Al-Tirmzi, and the compilation of Al-Albani, and the accuracy of the proposed approach was 97.60%.

Harrag et al. [18] proposed an association rule-based approach to build Hadith’s ontology from Sahih Al-Bukhari’s book, the study did not mention how many Hadiths were used, and also, the study did not remark any results. In [19], the researchers introduced a graph construction for narration series depending on k-nearest neighbor, Naïve Bayes, and decision tree. The experiments were conducted on Hadiths from Sahih Al-Bukhari’s book and Musnad Ibn Hanbal book, the study did not mention how many Hadiths were used, and the F-measures were 85%, 80%, and 86% for the k-nearest neighbor, Naïve Bayes, and decision tree, respectively.

Azmi and AIOfaidly [20] introduced a Heuristic rule-based approach to classifying the authenticity of Hadith, the experiments were conducted on 752 Hadiths from Sunan Al-Tirmzi and 2,180 Hadiths from Sahih Al-Bukhari, and the success rates were 93.6% and 99.6% for Sunan Al-Tirmzi and Sahih Al-Bukhari, respectively. Harrag [21] presented a text mining approach for knowledge extraction in Hadith, the approach was built using finite-state transducers (FSTs), the experiments were conducted on the complete set of Sahih Al-Bukhari’s Hadiths, and the F-measure was 52%. Abd Rahman et al. [22] proposed a system for recognizing narrators’ names in the Malay language depending on the rule-based approach, the experiments were conducted on 150 Hadiths from Sahih Al-Bukhari, and the study did not remark any results.

Alhawarat [23] introduced a technique that combines the rule-based method with the n-grams model to elicit narrators’ names from Hadiths, the experiments were conducted on the six books of Hadith, and the F-measures were 65.11% for the n-grams model and 70.76% for the n-grams and the rule-based approach. Faidei et al. [24] combined numerous classifiers and stemmers to compare Hadith classification tools, the experiments were conducted on 795 Hadiths from Sahih Al-Bukhari, and the best combination was the stemmer of Khoja with SVM, which got 57.50% as accuracy. Balgasem and Zakaria [25] used the log-likelihood ratio (LLR) with the rule-based approach to identifying narrators’ names, the experiments were conducted on 235 Hadiths from Sahih Al-Bukhari, and the F-measure was 82%.

Najib et al. [26] proposed a system to classify Hadiths (in the Malay language) based on k-nearest neighbor (k-NN), Naïve Bayes (NB), and support-vector machine (SVM), the experiments were conducted on 50 Hadiths from Sunan Al-Tirmzi and 50 Hadiths from Sahih Al-Bukhari, and the accuracies were 62%, 81%, and 82% for k-NN, NB, and SVM, respectively. Mahmood et al. [27] constructed an IR system depending on the conditional random field (CRF) and finite-state transducers (FSTs), the experiments were conducted on 7,563 Hadiths (in the Urdu language) from Sahih Al-Bukhari, and the F-measure was 92.41%.

Sari et al. [28] proposed an HMM-based method for treating Hadiths written in the Indonesian language, and the indexed data contain the person’s name, Hadiths’ collections, and Hadiths’ numbers. The experiments were conducted on 38,102 Hadiths from the books of Tirmzi, Nasai, Malik, Ibnu Majah, Darimi, Ahmad, Al-Bukhari, Muslim, and Abu Dawud, and the F-measure was 86%. Najeeb [4] suggested a new method to process Hadiths built on genetic algorithms (GAs), and this method seeks to foretell narrators’ names of Hadiths and the rest part of Isnads (POIs) as well. The experiments were conducted on 3,033 from the Sahih Muslim book, and the accuracy was 81.44%.

In [29], Najeeb et al. present the actions to launching an ANLP scientific laboratory in Al-Qunfudah college, Umm Al-Qura University, KSA, to serve Arabic and Islamic research studies. The same authors built a corpus-based lexicon to serve Hadith science [30]. An XML database for narrators and Hadiths was introduced in [31]. Saloot et al. [32] carried out comparative research of data mining and classification techniques of Hadiths.

In [33], the authors presented a regular expression-based approach to address web pages that comprise Hadiths, and the researchers built a conventional database by extracting data from Hadiths. Abdelaal et al. [34] used data mining and machine learning techniques to propose a classification system that forecasts the topic of Hadith such as fasting and prayer. Najeeb [35] inspects the field of employing deep learning in Hadith science. Fadele et al. [36] examined various detection techniques of the bogus Hadiths, and this includes data-based and knowledge-based techniques.

Saeed et al. [37] introduced a social network analysis for narrators, they discussed the structure of these networks and their properties and also examined the interaction patterns and the dominant narrators of these networks, and the authors suggested a new raking approach for narrators. Ubishat et al. [38] introduced a new classification approach for Hadiths based on the sine cosine algorithm (SCA). To improve the SCA exploitation, the simulated annealing (SA) algorithm was used, and to improve solution diversity, the singer chaotic map was used. Plentiful websites [39–45] provide various Hadith services such as e-books, translation tools, and search tools.

In [46], the author disseminated an important review of the scientific methods regarding Hadiths in the following types: natural language processing (NLP), information retrieval (IR), and knowledge extraction (KE). The researcher explained the difficulty of doing a comparison between diverse research studies related to Hadith science for two reasons: first, there are varied preprocessing steps used by these research studies, and second, there are varied Hadiths used by these research studies as a corpus.

Azmi et al. [47] introduced a remarkable review on Hadith research studies, and they investigate the computational and natural language processing research studies and classified them into three fields: Hadith content-based research studies, narration-based research studies, and overall research studies. Binbeshir et al. [48] conducted a systematic review on Hadith authentication and classification methods, and they compared 27 types of research: 13

types of research in classification and 14 types of research in the authentication. All classification research studies used Matn, while most of the authentication research studies used Isnad. The authors pointed out the shortage of exploring some feature extractions, like syntactic and semantic.

Table 3 presents a comparison of research studies related to Hadith processing: the research studies ordered by time of publication.

As it can be observed, various research studies proposed numerous techniques to identify narrators' names; however, there are no studies to handle Isnads of Hadith using HMM and gazetteer lists for Arabic Hadiths. This study presents a new hybrid approach based on the hidden Markov model (HMM) and gazetteer method to process Arabic Isnads. The approach objective is to expect the narrators' names and the other POIs in the Isnad. This technique tags the Isnad-Phrases with predefined POIs: Prophet-Name, Narrator-Name, Narrator-Name-Prefix, Received-Method, Received-Method-Prefix, Replacement, Title, and Others.

3. The Proposed Approach

The main architecture of the proposed hybrid approach is shown in Figure 1; this approach combines the gazetteer lists with the HMM method. The proposed approach deals with 1,000 Hadiths from "Sahih Muslim": 900 Hadiths as a training dataset and 100 Hadiths as a testing dataset. The testing dataset represents the raw Hadiths, which are considered as the input for this approach. The "preprocessing phase" will treat these raw Hadiths—as explained later—then, the gazetteer lists will be invoked to predict the POI categories for Isnad-Phrases of all Hadiths. Subsequently, the HMM method tries to predict the most proper POI categories for all Isnad-Phrases that not predicted by the gazetteer lists. The results will be the "recognized POI" for all Isnad-Phrases in all Hadiths.

The preprocessing phase is shown in Figure 2. The raw Hadiths will be divided into "Matn" and "Isnad"; then, the Isnad-Phrase extraction process will be conducted by an expert scientist in "Hadith science." Subsequently, all Isnad-Phrases will be assigned to the predefined POI categories. Finally, the Hadith-annotated file will be constructed by all these Isnad-Phrases and its equivalent POIs.

Gazetteer lists accept the Hadith-annotated file as a lookup file; when a new Isnad is coming, the gazetteer lists look for its Isnad-Phrases in the Hadith-annotated file and assigns them to the proper POIs. Note that not all Isnad-Phrases will be assigned in this method because some of these Isnad-Phrases may not exist in the Hadith-annotated file. Also, some of the predictions will not be correct due to the ambiguity problem.

HMM uses the Hadith-annotated file for training purposes; subsequently, the Viterbi algorithm is used to get the final POI results. In the training phase, HMM accepts the Hadith-annotated file as input and uses it to determine the following 4 HMM parameters: States-Array, Start-Probability (π) array, transition probability matrix (A), and emission probability matrix (B). To determine the States-

Array, all the POIs are checked in the Hadith-annotated file and then inserted in the States-Array without duplication.

Algorithm 1 explains the process of finding the Start-Probability (π) array. Start-Probability (π) array represents the probability of starting one Isnad with a certain POI. For every POI, the start probability of this POI will equal the number of all Isnads in the Hadith-annotated file that starts with this POI divided by the total number of Isnads in the Hadith-annotated file, as shown from line 5 to line 8.

Algorithm 2 explains the process of finding the transition probability matrix (A), and it aims to calculate the transition probabilities between states. The input file of this algorithm is the Hadith-annotated file, while the output is the matrix of transition probability (A). To calculate the probability of transition from state S_i to state S_j (i.e., the probability of the sequence $S_i S_j$), the algorithm calculates the number of $S_i S_j$ sequence occurrences in the Hadith-annotated file and then divides this number by the total number of S_i occurrences in the Hadith-annotated file, as shown from line 7 to 9.

Algorithm 3 shows the process of finding the emission probability (B) matrix; this process aims to calculate the probability of assigning an Isnad-Phrase to a specific POI. The algorithm checks all unique Isnad-Phrases in the Hadith-annotated file and calculates the emission probability for each of them according to a specific POI. So, the emission probability is calculated by dividing the number of times that a certain Isnad-Phrase assigns to a certain POI by the total number of occurrences of this POI in the Hadith-annotated file, as shown from line 6 to line 9.

The following example shows the implementation of the proposed approach using two Hadiths from the "Sahih Muslim" book, these Hadiths are listed in Table 4. The POIs of the first and second Hadiths are shown in Tables 5 and 6 respectively. The combination of these tables formulates the Hadith-annotated file for this example.

The HMM parameters can be calculated as follows:

- (1) States-Array: States-Array represents all POIs in the Hadith-annotated file without duplication. Therefore, the States-Array for this example is as follows: {Received-Method, Narrator-Name-Prefix, Narrator-Name, Received-Method-Prefix, Prophet-Name, Others}.
- (2) Start-Probability (π) Array: The Start-Probability (π) array for this example could be calculated as explained in Algorithm 1, and the results are shown in Table 7.
- (3) Transition probability matrix (A): The transition probability matrix (A) for this example could be calculated as explained in Algorithm 2, and the results are shown in Table 8.
- (4) Emission probability matrix (B): The emission probability matrix (B) for this example can be formed as explained in Algorithm 3, and the results are shown in Table 9. Note that the Hadith-annotated file contains many Isnad-Phrases that could not be displayed in this table, consequently, just a

TABLE 3: Comparison of Hadith processing research studies.

Research	Scope of research	Source of data	Results
Harrag and Hamdi-Cherif [10]	Classification of Hadiths depending on the query of the user. Vector space model, cosine, TF-IDF are used in the study.	60 Hadiths	Precision = 66% Recall = 80%
Ghazizadeh et al. [11]	Classification of Hadith authenticity depending on fuzzy technique	Al-Kafi book, but they did not mention how many Hadiths they used in their study	Accuracy = 94%
Harrag and El-Qawasmah [12]	Classification of Hadiths depending on the artificial neural network (ANN) and singular value decomposition (SVD)	453 Hadiths	F-measures: ANN = 85.75% ANN + SVD = 88.33%
Bounhas et al. [13]	Naïve Bayes (NB) classifier for narration chain reliability	1000 Hadiths	F-measure = 89.01%
Azmi and Bin Badia [14]	Graph construction for narration series depending on context-free grammar (CFG) and semantic web ontology	90 Hadiths from Sahih Muslim and Sahih Al-Bukhari	Success rate = 86.7%
Azmi and Bin Badia [15], Azmi and AlBadia [16]	Graph construction for narration series depending on context-free grammar (CFG) and memory-based learning	90 Hadiths from Sahih Muslim and Sahih Al-Bukhari	Success rate = 86.7%
Aldhahm et al. [17]	Decision tree (DT) approach to classify the authenticity of hadiths	999 Hadiths from Sahih Al-Bukhari, Sunan Al-Tirmzi, and the compilation of Al-Albani	Accuracy = 97.60%
Harrag et al. [18]	Association rule-based approach to build Hadith's ontology	Sahih Al-Bukhari book, but they did not mention how many Hadiths they used in their study	The study did not mention any results
Siddiqui et al. [19]	Graph construction for narration series by the k-nearest neighbor, Naïve Bayes, and decision tree	Sahih Al-Bukhari book and Musnad Ibn Hanbal book, but they did not mention how many Hadiths they used in their study	F-measures: k-nearest neighbor = 85% Naïve Bayes = 80%, Decision tree = 86%
Azmi and AlOfaidly [20]	Heuristic rule-based approach to classify the authenticity of Hadiths	752 from Sunan Al-Tirmzi 2,180 Hadiths from Sahih Al-Bukhari	Success rate: Sunan Al-Tirmzi = 93.6% Sahih Al-Bukhari = 99.6%
Harrag [21]	Information extractor using FST	Sahih Al-Bukhari: complete set of Hadiths	F-measure = 52%
Abd Rahman et al. [22]	Recognizing narrators' names depending on rule-based approach	150 Hadiths from Sahih Al-Bukhari (in the Malay language)	The study did not mention any results
Alhawarat [23]	Extraction of narrators depending on rules and n -grams model	The six books of Hadith	F-measures: n -grams model = 65.11% n -grams and rule-based approach = 70.76
Faidi et al. [24]	Combine numerous classifiers and stemmers to compare Hadith classification tools	795 Hadiths from Sahih Al-Bukhari	The best accuracy is 57.50% (stemmer of Khoja combined with SVM)
Balgasem and Zakaria [25]	Using log-likelihood ratio (LLR) with rule-based approach to identifying narrators' names	235 Hadiths from Sahih Al-Bukhari	F-measure = 82%
Najib et al. [26]	Classification of Hadiths (in the Malay language) using k-nearest neighbor (k-NN), Naïve Bayes (NB), and support vector machine (SVM)	50 Hadiths from Sunan Al-Tirmzi and 50 Hadiths from Sahih Al-Bukhari (in the Malay language)	Accuracy: K-NN = 62% NB = 81%, SVM = 82%,
Mahmood et al. [27]	Construction of IR system depending on the conditional random field (CRF) and finite-state transducers (FSTs)	7,563 Hadiths (in the Urdu language) from Sahih Al-Bukhari	F-measure = 92.41%
Sari et al. [28]	Hidden Markov model approach to index the names in Hadiths	38,102 Hadiths (in Indonesian language) from the books of Tirmzi, Nasai, Malik, Ibnu Majah, Darimi, Ahmad, Al-Bukhari, Muslim, and Abu Dawud	F-measure = 86%
Najeeb [4]	Predicting narrators' names and the other Isnad parts using genetic algorithm (GA)	3,033 from Sahih Muslim book	Accuracy = 81.44%.

FIGURE 1: Architecture of the proposed hybrid approach.

FIGURE 2: Preprocessing phase.

```

(1) Function Name: FindStartProbabilityArray
(2) Function Input: Hadith-Annotated file
(3) Function Output: Start-Probability-Array
(4) START
(5)   FOR each POI in Hadith-Annotated file
(6)     A = Count Isnads which start with this POI
(7)     B = Total number of Isnads in Hadith-Annotated file
(8)     Start-Probability = (A/B)
(9)   END
  
```

ALGORITHM 1: Find Start Probability (π) Array.

```

(1) Function Name: FindTransitionProbabilityMatrix
(2) Function Input: Hadith-Annotated file
(3) Function Output: TransitionProbabilityMatrix
(4) START
(5)   FOR every state ( $S_i$ ) in States-Array
(6)     FOR every state ( $S_j$ ) in States-Array
(7)       W = Number of ( $S_i S_j$ ) sequence occurrences in the Hadith-Annotated file
(8)       Z = Total number of  $S_i$  occurrences in Hadith-Annotated file
(9)       Transition-Probability = (W/Z)
(10)   END
  
```

ALGORITHM 2: Find Transition Probability (A) Matrix.

```

(1) Function Name: FindEmissionProbabilityMatrix
(2) Function Input: Hadith-Annotated file
(3) Function Output: EmissionProbabilityMatrix
(4) START
(5)   FOR every unique Isnad-Phrase within Hadith-Annotated file
(6)     FOR every unique POI within Hadith-Annotated file
(7)       N = Count how often the Isnad-Phrase assigns to this POI
(8)       T = Total occurrences number of this POI in Hadith-Annotated file
(9)       Emission-Probability = (N/T)
(10)   END
  
```

ALGORITHM 3: Find Emission Probability (B) Matrix.

TABLE 7: Start probabilities (π).

Part of Isnad (POI)	Start probabilities
Received-Method	2/2
Narrator-Name-Prefix	0/2
Narrator-Name	0/2
Received-Method-Prefix	0/2
Prophet-Name	0/2
Others	0/2

TABLE 8: Transition probability matrix (A).

	Received-Method	Narrator-Name-Prefix	Narrator-Name	Received-Method-Prefix	Prophet-Name	Others	Title	Replacement
Received-Method	0/9	3/9	5/9	0/9	1/9	0/9	0/9	0/9
Narrator-Name-Prefix	0/3	0/3	3/3	0/3	0/3	0/3	0/3	0/3
Narrator-Name	5/9	0/9	1/9	2/9	0/9	1/9	0/9	0/9
Received-Method-Prefix	2/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
Prophet-Name	0/2	0/2	0/2	0/2	0/2	1/2	0/2	0/2
Others	0/2	0/2	0/2	0/2	1/2	0/2	0/2	0/2

TABLE 9: Emission probability matrix (B).

...	بِئْسَ لَكَ اللَّهُ عَزَّ وَجَلَّ	مَعْتَسُ بْنُ قَالٍ	عَدِيَّ بْنَ حَاتِمٍ	عَبْدُ اللَّهِ بْنِ مَعْقِلٍ	إِسْحَاقُ أَبِي عَنَ			
...	0/9	1/9	0/9	0/9	0/9	0/9	0/9	6/9
...	0/3	0/3	0/3	0/3	0/3	0/3	3/3	0/3
...	0/9	0/9	0/9	1/9	1/9	0/9	0/9	0/9
...	0/2	0/2	1/2	0/2	0/2	0/2	0/2	0/2
...	1/2	0/2	0/2	0/2	0/2	0/2	0/2	0/2
...	0/2	0/2	1/2	0/2	0/2	0/2	0/2	0/2

TABLE 10: Accuracy of gazetteer lists.

	Received-Method	Narrator-Name-Prefix	Narrator-Name	Received-Method-Prefix	Prophet-Name	Others	Title	Replacement
Number of Isnad-Phrases	714	348	670	152	100	279	26	49
Correct	528	251	418	117	96	244	17	33
Accuracy	74%	72%	62%	77%	96%	87%	65%	67%

TABLE 11: Average Accuracy for all methods.

	Gazetteer lists (%)	HMM method (%)	The proposed hybrid approach (%)
Average accuracy	75	83	86

HMM method has better accuracy than gazetteer lists. The last experiment was conducted on the same Hadiths using the proposed hybrid approach, which combines gazetteer lists with the HMM method, and the results are expressed in Table 13. As explained in Table 11, the average accuracy of all POI predictions using the proposed hybrid approach is 86%. Table 11 compares the average accuracy between these three methods.

The results show that the proposed hybrid approach achieves higher accuracy among the other methods because gazetteer lists will be invoked to predict the POI categories for Isnad-Phrases of all Hadiths; then, the HMM method tries to predict the most proper POI categories for all Isnad-Phrases that are not predicted by gazetteer lists; so, this technique will increase the accuracy of prediction.

TABLE 12: Accuracy of HMM method.

	Received-Method	Narrator-Name-Prefix	Narrator-Name	Received-Method-Prefix	Prophet-Name	Others	Title	Replacement
Number of Isnad-Phrases	714	348	670	152	100	279	26	49
Correct	589	301	594	123	94	216	19	41
Accuracy	82%	86%	89%	81%	94%	77%	73%	84%

TABLE 13: Accuracy of the proposed hybrid method.

	Received-Method	Narrator-Name-Prefix	Narrator-Name	Received-Method-Prefix	Prophet-Name	Others	Title	Replacement
Number of Isnad-Phrases	714	348	670	152	100	279	26	49
Correct	605	312	618	126	96	244	19	41
Accuracy	85%	90%	92%	83%	96%	87%	73%	84%

5. Conclusions

A new hybrid approach to process Isnads of Hadiths was introduced in this study. The proposed approach, which is based on hidden Markov model (HMM) and gazetteer lists, is used to expect all part of Isnads (POIs) in the Isnad including narrators' names. The experiments were conducted on 1,000 Hadiths from "Sahih Muslim": 900 Hadiths as training dataset and 100 Hadiths as testing dataset. The results show that the proposed hybrid approach achieves better accuracy than HMM or gazetteer lists. The results show that the accuracy of the proposed approach is 86%, which is lower than the desired objective; consequently, future works aim to enhance the training process of the HMM using genetic algorithms (GAs). Subsequently, the enhanced method will be used to introduce a novel "judgment" approach for Hadiths, which will differentiate the trustworthy (correct) Hadiths from the bogus (false) Hadiths.

Data Availability

The Hadiths of "Sahih Muslim" in the Arabic language are available from many websites (<http://almeshkat.net/book/63>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The author would like to thank the Deanship of Scientific Research at Umm Al-Qura University for the financial support under grant no. 43508026. As well, the author would like to express his sincere thanks and gratitude to Dr. Naser Younes Sabra, Ph.D., in Islamic Studies and Hadith sciences, for his valuable guidance and support; without him, this research would not have been completed.

References

- [1] M. Al-Tahhan, *Tisir Mustalah al Hadith* [Clarification of Hadith's Terminology] (in Arabic), Maarif Library for Publishing and Distribution, KSA, 2011.
- [2] M. M. Al-Azami, *Studies in hadith Methodology and Literature*, American Trust Publication, Indianapolis, IN, USA, 1978.
- [3] M. Al-Hajjaj, *Sahih Muslim* [Muslim Probe] (in Arabic), Dar Ibn Al-Jawzi Publication and Distributors, Egypt, 2009.
- [4] M. M. A. Najeeb, "A novel hadith processing approach based on genetic algorithms," *IEEE Access*, vol. 8, pp. 20233–20244, 2020.
- [5] X. Wu, Z. Wu, J. Jia, and L. Cai, "Adaptive Named Entity Recognition Based on Conditional Random fields with Automatic Updated Dynamic Gazetteers," in *Proceedings of the 8th International Symposium On Chinese Spoken Language Processing*, pp. 363–367, Hong Kong, China, December 2012.
- [6] N. Zamin and A. Oxley, Edited by J. M. Zain, Ed., "Building a Corpus-Derived Gazetteer for Named Entity Recognition," in *Software Engineering And Computer Systems ICSECS, Communications In Computer and Information Science*, W. M. Mohd and E. El-Qawasmeh, Eds., Springer, Berlin, Germany, 2011.
- [7] V. Sharan, S. Kakade, P. Liang, and G. Valiant, "Prediction with a short memory," in *Proceedings of the 50th annual ACM SIGACT Symposium on Theory of computing*, pp. 1074–1087, Los Angeles CA, USA, 2018.
- [8] J. Yang, B. Song, B. Yan, and G. Li, "A novel hidden Markov model for genome-wide association studies," in *Proceedings of the 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 500–506, Prague, July 2017.
- [9] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [10] F. Harrag and A. Hamdi-Cherif, "UML modeling of text mining in Arabic language and application to the prophetic traditions "Hadiths"" in *Proceedings of the First International Symposium on Computers and Arabic Language and Exhibition*, pp. 11–20, Riyadh, Saudi Arabia, November 2007.
- [11] M. Ghazizadeh, M. H. Zahedi, M. Kahani, and B. M. Bidgoli, "Fuzzy expert system in determining hadith validity," in *Advances In Computer and Information Sciences and Engineering*, T. Sobh, Ed., Springer, Berlin, Germany, pp. 354–359, 2008.
- [12] F. Harrag and E. El-Qawasmah, "Neural network for Arabic text classification," in *Proceedings of the Second International*

- conference on the applications of digital information and web technologies, pp. 778–783, Icaditw, UK, 2009.
- [13] I. Bounhas, B. Elayeb, F. Evrard, and Y. Slimani, "Toward a computer study of the reliability of Arabic stories," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 8, pp. 1686–1705, 2010.
 - [14] A. M. Azmi and N. Bin Badia, "E-Narrator-an application for creating an ontology of Hadiths narration tree semantically and graphically," *The Arabian Journal for Science and Engineering*, vol. 35, no. 2, pp. 51–68, 2010.
 - [15] A. M. 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 '10)*, pp. 1–7, Beijing, China, August 2010.
 - [16] A. Azmi and N. AlBadia, "Mining and visualizing the narration tree of hadiths (prophetic traditions)," in *Proceedings of the Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches*, pp. 239–257, Hershey, PA, USA, 2012.
 - [17] K. Aldhaln, A. Zeki, A. Zeki, and H. Alreshidi, "Improving knowledge extraction of Hadith classifier using decision tree algorithm," in *International Conference on Information Retrieval & Knowledge Management*, pp. 148–152, Malaysia, 2012.
 - [18] F. Harrag, A. Alothaim, A. Abanmy, F. Alomaigan, and S. Alsalehi, "Ontology extraction approach for prophetic narration (hadith) using association rules," *International Journal on Islamic Applications in Computer Science and Technology*, vol. 1, no. 2, pp. 48–57, 2013.
 - [19] M. A. Siddiqui, M. 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.
 - [20] A. M. Azmi and A. M. AlOfaidly, "A novel method to automatically pass HUKM on hadith," in *Proceedings of the 5th International Conference on Arabic Language Processing (CITALA '14)*, pp. 118–124, Qujda, Morocco, November 2014.
 - [21] F. Harrag, "Text mining approach for knowledge extraction in Sahih Al-Bukhari," *Computers in Human Behavior*, vol. 30, pp. 558–566, 2014.
 - [22] N. Abd Rahman, N. Alias, N. K. Ismail, Z. b. Mohamed, and M. N. Alias, "An identification of authentic narrator's name features in Malay hadith texts," in *Proceedings of the 2015 IEEE Conference on Open Systems (ICOS)*, pp. 79–84, Malaysia, August 2015.
 - [23] 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.
 - [24] K. Faidi, R. Ayed, I. Bounhas, and B. Elayeb, "Comparing Arabic NLP tools for hadith classification," *International Journal On Islamic Applications In Computer Science & Technology*, vol. 3, no. 3, pp. 1–12, 2015.
 - [25] S. Balgasem and L. Zakaria, "A hybrid method of rule-based approach and statistical measures for recognizing narrators name in hadith," in *Proceedings of the 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, pp. 1–5, Malaysia, November 2017.
 - [26] S. R. Najib, N. Abd Rahman, N. K. Ismail, and N. Alias, Z. b. Mohamed Nor and M. N. Alias, Comparative study of machine learning approach on Malay translated hadith text classification based on sanad," in *Proceedings of the Eighth International Conference on Mechanical and Manufacturing Engineering (ICME'17)*, Malaysia, 2017.
 - [27] A. Mahmood, H. U. Khan, Z. Rehman, and W. Khan, "Query based information retrieval and knowledge extraction using Hadith datasets," in *Proceedings of the 2017 13th International Conference on Emerging Technologies (ICET)*, pp. 1–6, Islamabad, December 2017.
 - [28] W. Sari, M. Arif Bijaksana, and A. Huda, "Indexing Name in Hadith Translation Using Hidden Markov Model (HMM)," in *Proceedings of the 7th International Conference On Information And Communication Technology (ICoICT)*, pp. 1–5, Kuala Lumpur, Malaysia, July 2019.
 - [29] M. Najeeb, A. Abdelkader, and M. Al-Zghoul, "Arabic natural language processing laboratory serving Islamic sciences," *International Journal of Advance Computer Science Applications*, vol. 5, no. 3, pp. 114–117, 2014.
 - [30] M. Najeeb, A. Abdelkader, M. Al-Zghoul, and A. Osman, "A lexicon for hadith science based on a corpus," *International Journal of Advance Computer Science Applications*, vol. 6, no. 2, pp. 1336–1340, 2015.
 - [31] M. A. Najeeb, "XML database for hadith and narrators," *American Journal of Applied Sciences*, vol. 13, no. 1, pp. 55–63, 2016.
 - [32] M. A. Saloot, N. Idris, R. Mahmud, S. Ja'afar, D. Thorleuchter, and A. Gani, "Hadith data mining and classification: a comparative analysis," *Artificial Intelligence Review*, vol. 46, no. 1, pp. 113–128, 2016.
 - [33] A. Mahmood, H. Khan, F. Alarfaj, M. Ramzan, and M. Ilyas, "A multilingual datasets repository of the hadith content," *International Journal of Advanced Computer Science and Applications*, vol. 92, pp. 165–172, 2018.
 - [34] H. M. Abdelaal, B. R. Elemery, and H. A. Youness, "Classification of hadith according to its content based on supervised learning algorithms," *IEEE Access*, vol. 7, Article ID 152387, 2019.
 - [35] 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.
 - [36] A. A. Fadele, A. Kamsin, K. Ahmad, and H. Habiba, "A novel classification to categorise original hadith detection techniques," *International Journal of Information Technology*, Springer, 2021.
 - [37] S. Saeed, S. Yousuf, F. Khan, and Q. Rajput, "Social network analysis of Hadith narrators," *Journal of King Saud University-Computer and Information Sciences*, 2021.
 - [38] M. Tubishat, S. Ja'afar, N. Idris, and M. A. Al-Betar, "Improved sine cosine algorithm with simulated annealing and singer chaotic map for Hadith classification," *Neural Computing & Applications*, vol. 34, pp. 1385–1406, 2021.
 - [39] Hadith, "Hadith library," 2021, <https://ahadith.co.uk/>.
 - [40] Shamela, "المكتبة الشاملة على الشبكة", <http://shamela.ws>.
 - [41] Dorar, "موقع الدرر السنية على الشبكة", <http://www.dorar.net>.
 - [42] Islamweb, "موقع إسلام ويب على الشبكة", <http://www.islamweb.net>.
 - [43] Sonnaonline, "جامع للحدیث النبوی", <http://www.sonnaonline.com>.
 - [44] Sunnah, "Sayings and teachings of Prophet Muhammad," 2021, <https://sunnah.com>.
 - [45] S. Alifta, "جامع خادم الحرمين الشريفين للسنة النبوية", 2021, <https://sunnah.alifta.gov.sa>.
 - [46] I. Bounhas, "On the usage of a classical Arabic corpus as a language resource: related research and key challenges," *ACM*

- Transactions on Asian and low-resource language information processing (TALLIP)*, vol. 18, no. 3, 2019.
- [47] A. M. Azmi, A. O. Al-Qabbany, and A. Hussain, "Computational and natural language processing based studies of hadith literature: a survey," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1369–1414, 2019.
 - [48] F. Binbeshr, A. Kamsin, and M. Mohammed, "A systematic review on hadith authentication and classification methods," *ACM transactions on Asian low-resource Language Information Processing*, vol. 34, p. 17, 2021.