IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 12, 2020, accepted January 16, 2020, date of publication January 21, 2020, date of current version January 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968417

A Novel Hadith Processing Approach Based on Genetic Algorithms

MOATH MUSTAFA AHMAD NAJEEB
Computer Science Department, College of Computing at Al-Qunfudah, Umm Al-Qura University, Mecca 21955, Saudi Arabia
e-mail: mmnajeeb@uqu.edu.sa

This work was supported by the Umm Al-Qura University, Deanship of Scientific Research, under Grant 15-COM-3-1-0003.

ABSTRACT Quran and Hadith are the main religious sources in Islam. Hadith is mainly the saying of Prophet Mohammad, and it consists of two parts: first is ''Isnad''; which is the series of persons (narrators) who report (narrate) Hadith, and the second is ''Matn''; which is the saying or the narration itself. Isnad can be divided into phrases called ''Isnad-Phrases'', which could be tagged according to proposed Part-Of-Isnads (POIs) that represent the entity types of Isnad, such as Narrator Name, Prophet Name and Received Method. One of the main objectives of Hadith sciences is to ascertain the validity of Hadith and determine if it is accepted or rejected; this is known as Hadith judgment. Hadith scholars put many rules to judge Hadiths; some of these rules are related to the chain of narrators. Therefore, to judge Hadith, first, the narrators' names in Isnad should be extracted then judgment rules can be applied. Many researches proposed different methods to extract narrators' names from Isnad. To the best of the author's knowledge, there is no research process Isnad using Genetic Algorithms (GA). In this research, a novel approach is introduced for Isnad processing based on GA. This approach aims to predict the narrators' names and the other POIs for Isnad. The experiments, which conducted on all Hadiths narrated in ''Sahih Muslim'' book, show that the proposed approach achieved 81.44% accuracy. For further research in this genre, investigation of utilizing other approaches for Isnad processing, such as deep learning, is recommended.

INDEX TERMS Arabic natural language processing, genetic algorithms, hadith, isnad, tagging.

I. INTRODUCTION

Hadith is considered as the second legalization source in Islam after Quran. It includes all the sayings and actions of Prophet Mohammad. After the death of Prophet Mohammad, his companions and the successors (who came after companions and received Hadiths from them) started collecting Prophet's Hadiths. For this, they studied Isnad, which represents the transmission chain that consists of many narrators. They set rules to distinguish between authentic and fake narrations, and so, various Hadith sciences had been established like ''Mustalah Al-Hadith'' and ''Al-Jarh Wa Al-Ta'dil''. These efforts started from the first century AH (seventh century CE) and continued for many. Hadith scholars set a lot of rules and principles related to Hadith science and wrote thousands of books and dictionaries to serve these sciences [1].

One of the essential sciences in Hadith is ''Al-Jarh Wa Al-Ta'dil'', it gives a methodological framework to investigate all the life details of the narrators. It also investigates, analyzes and evaluates all the narrators' aspects. Hence, it represents a vital tool for Hadiths scholars to assess Hadiths, so they need to utilize this science to examine the status of each narrator and then judge the Isnad.

''Al-Jarh Wa Al-Ta'dil'' consists of two main branches; first ''Al-Jarh'', which concerns with criticism of the narrator since he maybe a liar, unknown person, fake or has a bad memory. This criticism implies the investigation of the narrator's honesty, trustiness, sincerity, memory, etc. The narrators under this branch are divided into six ''Marateb'', i.e., ranks or levels. The first rank indicates that the narrator has the minimum criticism or wound in his trustiness or memorization. Many terms used to express this rank like: not safe, his Hadith is soft, etc. While, the sixth rank indicates that the narrator is entirely liar, many terms used to express this rank as he lies, he is fabricator, compulsive liar, etc. The second branch is ''Al-Ta'deel'', which represents the opposite case of ''Al-Jarh'' and concerns with situations that indicate praising of the narrator. It implies some expressions like reliable, proficient, unbiased, accurate, etc. These expressions show that

The associate editor coordinating the review of this manuscript and approving it for publication was Mustafa Servet Kiran.

the narrator is acceptable, and his narrations are trustworthy. The narrators under this branch are also divided into six ranks. Rank 1 indicates that the narrator has the highest level of praise and compliment, and many terms used to express this rank are; most established of the people, most reliable of the people, unmatched, etc. While rank 6 represents the lowest level of praise, and many terms used to express this rank are; satisfactory in Hadith, acceptable, etc.

Therefore, to judge the Isnad, Hadith scholars must have strong knowledge in ''Al-Jarh Wa Al-Ta'dil'' science and in syntax and semantics of Arabic language [2].

As explained earlier about parts of Hadith, ''Isnad'' and ''Matn. The following example (Hadith1) shows the Isnad and Matn for one Hadith from ''Sahih Muslim'' book [3]:

حَدَّثَنَا عُبَيْدُ اللهِ بْنُ سَعِيدٍ، وَعَبْدُ بْنُ حُمَيْدٍ، قَالَا: حَدَّثَنَا أَبُو عَامِرٍ الْعَقَدِيُّ، حَدَّثَنَا سُلَيْمَانُ بْنُ بِلَالٍ، عَنْ عَبْدِ اللهِ بْنِ دِينَارٍ، عَنْ أَبِي صَالِحٍ، عَنْ أَبِي هُرَيْرَةَ، عَنِ النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، قَالَ: ''الْإِيمَانُ بِضْعٌ وَسَبْعُونَ شُعْبَةً، وَالْحَيَاءُ شُعْبَةٌ مِنَ الْإِيمَانِ''.

Obaidullah bin Saeed narrated, and Abdo bin Humid, they said: Abu Amer Alaqdy narrated, Suleiman bin Bilal narrated, from Abdullah Ibn Dinar, from Abu Salih, from Abu Huraira that the Prophet peace be upon him said: "Iman (faith) has over seventy branches, and modesty is a branch of Iman".

The Isnad of Hadith1 is:

(حَدَّثَنَا عُبَيْدُ اللهِ بْنُ سَعِيدٍ، وَعَبْدُ بْنُ حُمَيْدٍ، قَالَا: حَدَّثَنَا أَبُو عَامِرٍ الْعَقَدِيُّ، حَدَّثَنَا سُلَيْمَانُ بْنُ بِلَالٍ، عَنْ عَبْدِ اللهِ بْنِ دِينَارٍ، عَنْ أَبِي صَالِحٍ، عَنْ أَبِي هُرَيْرَةَ، عَنِ النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، قَالَ)

(Obaidullah bin Saeed narrated, and Abdo bin Humid, they said: Abu Amer Alaqdy narrated, Suleiman bin Bilal narrated, from Abdullah Ibn Dinar, from Abu Salih, from Abu Huraira that the Prophet peace be upon him said)

And the ''Matn'' of Hadith1 is:

''الْإِيمَانُ بِضْعٌ وَسَبْعُونَ شُعْبَةً، وَالْحَيَاءُ شُعْبَةٌ مِنَ الْإِيمَانِ''

"Iman (faith) has over seventy branches, and modesty is a branch of Iman".

As Hadith scholars put rules and conditions to judge Hadiths, three of these conditions are specifically related to the transmission chain (list of narrators in the Isnad), these conditions are:

1- Continuity of the transmission chain, which means each narrator had received Hadith from his previous sheik (teacher), so the transmission chain is not broken.

2- The narrators have the characteristics of trustworthy; this implies two conditions, first: religiousness, which means avoidance of polytheism, sin, obscenity, etc. Second: morality or what we can call it a good personage; this means that the narrator is mentally sound, has good behavior, and acting appropriately regarding the traditions and norms of his community.

3- The narrators have the characteristics of preciseness; this implies two forms, first: preciseness in memory, which indicates the ability to recall Hadiths accurately, and the ability to approve that Hadith is correct if he hears it from someone else. Second: preciseness in writing, which indicates how the narrator writes Hadiths and how he recites it to other narrators.

So, if we aim to check these three conditions and ascertain the validity of Hadith, we must first predict the narrators' names from Isnad. Indeed, this prediction is a crucial step that precedes the judgment of Hadiths; therefore, we can summarize the importance of this work as follows:

1- If different Hadith scholars give more than one judgment on a single Hadith, we must have the ability to identify the correct one. So, we need to know the causes of this dispute and the preference factors used by scholars. The first step for this is to identify the narrators' names then applying the judgment rules.

2- Even if there is one judgment on a single Hadith, we must know how to ascertain the validity of this judgment, because it is not necessary that the scholar who judges this Hadith has a very high knowledge of judgment rules. From the first century AH (seventh century CE), hundreds of Hadith scholars investigated different Hadith sciences; however, the scholars are not at the same level of knowledge. There are scholars at the highest level of these sciences, such as Imam Ahmad, Al-Bukhari, Abu Hatem Al-Razi, Ali Ibn Al-Madini, and many scholars are below them in level. Hence, it is crucial to determine the authenticity of the Hadith's judgment.

3- Many narrations that represent the sayings of companions and successors have not been judged till now (not validated). These narrators have the same transmission chain structures as Hadiths, so, to ascertain the validity of these narrations, we need to identify the narrators' names then applying the judgment rules.

The structure of Hadith involves many types of entities so that it could be studied according to several aspects. Some researches in literature try to extract named entities from Hadiths books such as the number of chapters and sub-chapters, titles of chapters and sub-chapters, Isnads, Matns, version indications, etc. But, most of the researches are focused on the identification of narrators' names in Isnad since they represent the most important entities; however, Isnad contains more than the chain of narrators as it contains the Prophet name, received method, title, etc. Consequently, Part-Of-Isnad (POI) has been proposed in this research to indicate these entities, so Isnad can be divided into phrases called Isnad-Phrases. Each phrase can be assigned or tagged to one of the following Part-Of-Isnads (POIs): Prophet Name, Narrator Name, Narrator-Name Prefix, Received Method, Received-Method Prefix, Replacement, Title and Others. One Part-Of-Isnad (POI) represents one category that Isnad-Phrases can be assigned to it; table 1 shows the POIs of Hadith1.

Many researches have used various computational techniques to serve Hadith sciences; in this study, the researcher presents a novel approach called ''Isnad-GA'' to process Isnad using Genetic Algorithms (GA). This approach aims to predict the narrators' names and the other POIs for Isnad.

TABLE 1. Part-Of-Isnads (POIs) for Hadith1.

Isnad-Phrase/s	Part-Of-Isnad (POI)
حَدَّثَنَا (narrated)	Received Method
عُبَيْدُ اللهِ بْنُ سَعِيدٍ (Obaidullah bin Saeed)	Narrator Name
عَبْدُ بْنُ حُمَيْدٍ (Abdo bin Humid)	Narrator Name
قَالاَ (they said)	Received-Method Prefix
حَدَّثَنَا (narrated)	Received Method
أَبُو (Abu)	Narrator-Name Prefix
عَامِرٍ الْعَقَدِيُّ (Amer Alaqdy)	Narrator Name
حَدَّثَنَا (narrated)	Received Method
سُلَيْمَانُ بْنُ بِلَالٍ (Suleiman bin Bilal)	Narrator Name
عَنْ (from)	Received Method
عَبْدِ اللهِ بْنِ دِينَارٍ (Abdullah Ibn Dinar)	Narrator Name
عَنْ (from)	Received Method
أَبِي (Abu)	Narrator-Name Prefix
صَالِحٍ (Salih)	Narrator Name
عَنْ (from)	Received Method
أَبِي (Abu)	Narrator-Name Prefix
هُرَيْرَةَ (Huraira)	Narrator Name
عَنْ (from)	Received Method
النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ (Prophet peace be upon him)	Prophet name
قَالَ (said)	Others

GA is a stochastic algorithm that aims to discover the most appropriate solution in the search space. It derives its idea from genetic operators like selection, crossover and mutation, GA applied in different applications such as classification [4], bioinformatics [5] and handwriting recognition [6].

The remaining of this paper is organized as follows: Section II shows the related works, our approach is explained in Section III, the experimental results are discussed in Section IV, and the conclusion is presented in Section V.

II. RELATED WORKS

Hadith is the second source for legalization in Islam after the Holy Quran; therefore, various sciences introduced in Islamic eras to explain and serve Hadith. Many Muslim scholars set the foundation of these sciences; some of these sciences are ''Mustalah Al-Hadith'' and ''Al-Jarh Wa Al-Ta'dil''. The most reputed Hadith books that contain the ''Sahih'' (correct) Hadiths are: ''Sahih Al-Bukhari'', ''Sahih Muslim'', ''Sunan AlTermizi'', ''Sunan Abu Dawod'', ''Suana Ibn Majah'' and ''Sunan Al-Nasa'i'' [1].

In computing technology, significant efforts have been introduced to serve Hadith sciences; for example, some valuable websites [7]–[9] provide useful services related to Hadith. They give electronic versions for thousands of reputed Hadith books in Islamic heritage. Moreover, these websites offer many helpful tools for searching, translation, printing, etc.

Computer-science researches concerned with Hadith can be categorized into three major groups; first: researches deal with Hadith text, second: researches concentrate on narration chains, finally: miscellaneous researches. These researches are discussed below, and the empirical results are presented in Table 2 and Table 3.

A. HADITH-TEXT PROCESSING RESEARCHES

Harrag and Hamdi-Cherif [10] introduced a system called ''AuthenTique'' based on Vector Space Model. It uses text mining techniques to extract useful information from Hadiths and then compute the ''similarity'' of the user's query and Hadiths. The system retrieves a set of Hadiths in decreasing order of relevance of the query, the experiments conducted on 60 Hadiths only. The precision and recall of the system were 66% and 80% respectively. The same authors enhanced their work by publishing two more related researches [11], [12].

Harrag and El-Qawasmah [13] used Artificial Neural Networks (ANN) to classify Hadiths into fourteen categories. The researchers used three-layered back propagation architecture with singular values decomposition (SVD) as a preprocessing step. The experiments conducted on 435 Hadiths, the F-measure was 85% for ANN, and 88% for ANN with SVD.

Harrag *et al.* [14] created ontology based on Hadiths from Sahih Al-Bukhari; this ontology was related to Islamic jurisprudence (Fiqh). They used association rules, which considered as a significant data mining technique, to build this ontology. However, the researchers did not evaluate their proposed ontology.

In [15], Harrag introduced an entity-extraction method based on Finite State Transducers (FST) to extract useful entities from Hadiths of Sahih Al-Bukhari. This method detected a sequence of words containing relevant information from Hadiths. It also labeled the extracted entities according to a predefined set of labels, which are: Num-Kitab, Title-Kitab, Num-Bab, Title-Bab, Num-Hadith, Saned, Matn, Taalik, and Atraf. The method achieved 71% as precision, 39% as recall, and 52% as F-measure.

Faidi *et al.*[16] compared three text classification techniques for Hadith, namely: decision trees (DT), Naïve Bayes algorithm (NB), and Support Vector Machines (SVM), these techniques were implemented using WEKA toolkit. TF-IDF was used to evaluate the results of the classifiers. The researchers compared various stemming algorithms, such as Al-Stem Darwish, Al-Stem Alex, and Khoja's stemmer. The test dataset contains 795 Hadiths from Sahih Al-Bukhari, which are distributed among 23 categories. The experimental results showed that the accuracy of Khoja's stemmer and SVM combination was 57.50%; which outperformed the other stemmer and classifier combinations.

TABLE 2. Comparison of computer-science researches concerned with hadith text.

Reference	Research Domain & Approach	Data Source	Results
[10]	Hadiths classification based on similarity to user's query; using Vector Space Model, TF-IDF Weight, and Cosine measure	60 Hadiths	Precision = 66% Recall = 80%
[13]	Hadith classification using Artificial Neural Network (ANN) and Singular Value Decomposition (SVD)	453 diverse Hadiths from (The Encyclopedia of the Nine Books for the Honorable Prophetic Traditions, Harf Information Technology)	F-measure = 85.75% for ANN, and 88.33% for ANN+SVD
[14]	Ontology construction using Association Rules	An unspecified number of Hadiths from Sahih Al-Bukhari	No results reported
[15]	Finite State Transducer (FST) as information extractor	All Hadiths in Sahih Al-Bukhari	F-measure = 52%
[16]	Comparing Arabic NLP tools for Hadith classification using combinations of various stemmers and classifiers	795 Hadiths from Sahih Al-Bukhari distributed among 23 categories	Accuracy of Khoja's stemmer and SVM = 57.50% (best combination)
[17]	Evaluate various Machine Learning techniques in Hadith using one dataset	3150 Hadiths from Sahih Al-Bukhari	ANN accuracy = 94% (best method)
[19]	Information Retrieval system with Finite State Transducers (FST) and Conditional Random Field (CRF)	7563 Hadiths from Sahih Al-Bukhari (in the Urdu language)	F-measure = 92.41%

Saloot *et al.*[17] manage a notable comparative analysis of classification and data mining techniques that have been used for Hadiths. They divided their work into two tracks; one to investigate the researches that emphasis on classification process of Hadiths to some categories like authentic, not authentic, unknown, weak, etc. The other track investigates the researches that concentrate on using the data mining techniques to get useful information from Hadiths. Since these researches used different datasets of Hadith, the researchers decided to re-implement and re-evaluate many machine learning techniques using one dataset only. They used this technique because it is difficult to conduct a comparison between these researches while they are using different datasets. The single dataset that the researchers proposed contains 3150 Hadiths from Sahih Al-Bukhari. According to their experiments on the proposed dataset, the ANN used in [18] achieved the best accuracy, i.e. 94%.

Mahmood *et al.* [19] suggested an Information Retrieval system for Hadiths of Sahih Al-Bukhari. They used Finite State Transducers (FST) for entities extraction, and Conditional Random Field (CRF) as a Part of Speech (POS) tagger for Sanad and Matn. The tagger achieved 96.44% as precision, 88.77% as Recall, and 92.41% as F-Measure.

Table 2 shows a comparison of computer-science researches concerned with Hadith Text.

B. NARRATION-CHAINS PROCESSING RESEARCHES

Ghazizadeh *et al.* [20] presented a fuzzy system to give authenticity rates for Hadiths; these rates used to classify Hadiths into unknown, weak, goodness, reliable, and right. The researchers tested Hadiths from Al-Kafi book, and the approach achieved 94% accuracy.

Azmi and Bin Badia [21] developed a system called iTree, which displays the tree of narrators based on shallow parsing, memory-based learning, and Context-Free Grammar (CFG) that written in Extended Backus-Naur Form (EBNF). The experiments conducted on 90 Hadiths from Sahih Al-Bukhari and Sahih Muslim, the success rate of rendering the narration chain was 86.70%. The researchers provided more details about their system in [22]. The same researchers extend their previous study by applying semantic web ontology on Hadith; the new system was called "e-Narrator" [23].

Bounhas *et al.* [24] introduced a system that parsing the chain of narrators and recognizing the full names in it. It also checks the reliability of the narrators' chain according to three criteria, i.e., the credibility of the narrators, continuity of the chain, and reliability of the transmission. For the first two criteria, the system used metadata of narrators, which contains information about the narrators, and the relationships between narrators. For the last criterion, the system investigated the terms used in the Isnad of the Hadith. The experiments conducted on 1000 Hadiths using NB classifier; the F-measure of the identity recognition was 89.01%. The same researchers enhanced their study and presented generic guidelines for automatic reliability evaluation [25]. They also developed a mapping platform that implements interactive visualization mechanisms to understand the social path of information and reliability scores. They presented an illustrative case study to evaluate Hadiths automatically.

Aldhaln *et al.* [26] used decision trees for classification of Hadith validity; they classified Hadith into four categories: "Sahih", "Hasan", "Daief" and "Maudo". The researchers tested 999 Hadiths from Sahih Al-Bukhari, Sunan AlTermizi, and Al-Albani's compilation. The accuracy of the approach was 97.60%.

Azmi and AlOfaidly [27] introduced a heuristic rules-based approach for judgment of Hadiths. They used pure deterministic scheme that assigns a weight for each narrator

TABLE 3. Comparison of computer-science researches concerned with narration chains.

Reference	Research Domain / Approach	Data Source	Results
[20]	Classifying Hadith authenticity using Fuzzy expert system	An unspecified number of Hadiths from Al-Kafi book	Accuracy = 94%
[21], [22]	Represent narration chains as a graph using: shallow parsing, memory-based learning and Context-Free Grammar (CFG)	90 Hadiths from Sahih Al-Bukhari and Sahih Muslim	Success rate = 86.7%
[23]	Represent narration chains as a graph using: shallow parsing, memory-based learning, Context-Free Grammar (CFG) and semantic web ontology	90 Hadiths from Sahih Al-Bukhari and Sahih Muslim	Success rate = 86.7%
[24]	Narration chain reliability using Naïve Bayes (NB) classifier	1000 Hadiths - no source mentioned	F-measure = 89.01% (for Identity Recognition)
[26]	Classifying Hadith authenticity using Decision Trees (DT)	999 Hadiths from Sahih Al-Bukhari, Sunan AlTermizi, and Al-Albani's compilation	Accuracy = 97.60%
[27]	Judgment of Hadiths through classifying Hadiths authenticity using Heuristic rules	2180 Hadiths from Sahih Al-Bukhari and 752 from Sunan AlTermizi	Success rate = 99.6% for Sahih Al-Bukhari, and 93.6% for Sunan AlTermizi
[28]	Represent narration chains as a graph using: Naïve Bayes (NB), Decision Tree (DT), and k-Nearest Neighbor (k-NN)	An unspecified number of Hadiths from Sahih Al-Bukhari and Musnad Ibn Hanbal	F-measure = 80%, 86%, 85% for NB, DT and KNN respectively
[29]	Narrator extractions using n-grams model and rule-based approach	The Six books of Hadith ("Sahih Al-Bukhari", "Sahih Muslim", "Sunan AlTermizi", "Sunan Abu Dawod", "Suana Ibn Majah" and "Sunan Al-Nasa'i")	F-measure = 65.11% for n-grams model, F-measure = 70.76 for n-grams model integrated with Rule-based approach
[30]	A rule-based approach to recognize narrator names	445 narrators' names from 150 Hadiths from Sahih Al-Bukhari (in the Malay language)	No results reported
[31]	Classification of Malay translated Hadith based on their Isnad, using Support Vector Machine (SVM), Naïve Bayes (NB), and k-Nearest Neighbor (k-NN)	50 Hadiths from Sahih Al-Bukhari and 50 Hadiths from Sunan AlTermizi (in the Malay language)	Accuracy = 82%, 81%, 62% for SVM, NB and k-NN respectively
[32]	Recognizing narrators' names using rule-based approach with Log-Likelihood Ratio (LLR)	235 Hadiths from Sahih Al-Bukhari	F-measure = 82%

in the chain depends on three characters: trustworthiness, generation and deficiencies. At the end, the system grades Hadiths into three classes: Sahih, Hasan, and Dha'if. The researchers conducted their experiments on 2180 Hadiths from Sahih Al-Bukhari and 752 Hadiths from Sunan AlTermizi; the success rates for these books were 99.6% and 93.6% respectively.

Siddiqui *et al.* [28] used Named Entity Recognition to determine the narrators of Hadith; they displayed the narrators' chains visually as a network. The researchers manually tagged the tokens of Hadiths, and they selected the dataset from Sahih Al-Bukhari and Musnad Ibn Hanbal, however, they did not indicate the number of these Hadiths. The experiments conducted by testing three classifiers: Naïve Bayes (NB), Decision Tree (DT), and k-Nearest Neighbor (k-NN), the F-measures for these classifiers were 80%, 86% and 85% respectively.

Alhawarat [29] employed n-grams model to extract the narrators' names from Isnad. The names of narrators separated by phrases of narration, so the researcher collected a list of narrating terms which are the first words in the phrases of narration, then, an n-gram model was built from these phrases, the size of n-grams was from 3 to 15. The researcher also added a rule-based component to improve the results; this component used trigger words that appear before or after narrator names. The experiments conducted on the six books of Hadith ("Sahih Al-Bukhari", "Sahih Muslim", "Sunan AlTermizi", "Sunan Abu Dawod", "Suana Ibn Majah" and "Sunan Al-Nasa'i"), the F-measure was 65.11% for the n-gram model, and 70.76% for the n-gram model with the rule-based component.

In [30], the researchers introduced a rule-based method for recognizing the names of narrators in Malay text. They handled two problems; first, the absence of standard transliteration between Arabic and Malay languages for narrators' names; second, the different forms of the same narrator's name in Arabic. The experiments conducted on 445 narrators extracted from 150 Hadiths in Sahih Al-Bukhari, there are no results reported.

Najib *et al.* [31] introduced a system that classifies the Isnads of Hadiths in the Malay language into two categories: "Sahih Al-Bukhari" category and "Sunan AlTermizi" category. Three techniques were used to achieve this purpose, i.e., Support Vector Machine (SVM), Naïve Bayes (NB),

and k-Nearest Neighbor (k-NN). The researchers selected 50 Hadiths from Sahih Al-Bukhari and 50 Hadiths from Sunan AlTermizi to conduct the experiments; the SVM outperformed the other techniques and achieved 82% accuracy.

Balgasem and Zakaria [32] addressed a hybrid method for recognizing narrators' names in Hadith using rule-based and statistical methods. After the preprocessing phase, the researchers evaluated the rule-based approach using trigger words that represent the transmission terms; these terms often occur before narrators' names. On the other hand, the statistical approach compared Mutual Information (MI), Log-Likelihood Ratio (LLR), S-cost, R-cost, and U-cost. They evaluated the top N candidates ($N = 50$, 100, 150, and 200), LLR with $N = 200$ gave the best results. Therefore, the researchers combined the rule-based approach with LLR to enhance the results. The experiments conducted on 235 Hadiths from Sahih Al-Bukhari, and the F-measure of the method was 82%.

Table 3 shows a comparison of computer-science researches concerned with narration chains.

C. MISCELLANEOUS RESEARCHES

Bilal and Mohsin [33] introduced a new cloud system for Hadith classification called "Muhadith". It is an expert system that simulates Hadith scholars and aims to distinguish between authenticated and unauthenticated Hadiths. The cloud computing model is Software as a Service (SaaS), which is composed of an inference engine, knowledge base, parser, and explanation facility.

Several research laboratories originated to serve Arabic Natural Language Processing (ANLP) and Islamic sciences related to Quran and Hadith, for example, a specialized ANLP laboratory initiated at Umm Al-Qura University in KSA [34], some of the contributions of this lab can be found in [35], [36].

Mahmood *et al.* [37] introduced a multilingual datasets repository for Hadith. Their system extracts Hadith data in the form of a conventional database. The extraction was done from four books of Hadiths: Sahih Al-Bukhari, Sahih Muslim, Sunan Abu Dawod and Muwata' Malik, this database contains chapters, chains, narrators' names, and Matn of Hadiths. The researchers used Regular Expressions to treat web pages that contain Hadiths of these four books.

Bounhas [38] introduced a distinguished literature review of computer-science researches applied over Hadith. He studied and compared existing works related to Hadith in various fields of Natural Language Processing (NLP), Information Retrieval (IR), and Knowledge Extraction (KE). The researcher noticed that there is no synthetic work that summarizes the existing contributions and possible future directions. He also remarked that there are many publications in Hadith field did not reference to similar studies, so he established a reference for Hadith researchers who are working on comparative cases. This research pointed out that it is difficult to compare the various research works related to Hadith because they use different collections of Hadiths and several preprocessing steps. Therefore, the researcher discussed the possibility of building a generic Language Resource (LR) from Hadiths for open and free usage. He also recommended collaborative platforms to reduce the researchers' efforts and combining different tools to enhance tagging performance.

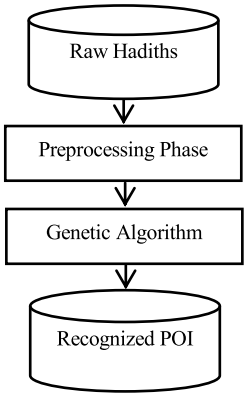

FIGURE 1. Architecture of the proposed approach.

Azmi *et al.* [39] presented a significant work that surveys the primary researches related to Hadith. They classified Hadith researches into three categories: Hadith content-based researches, narration-based researches, and overall researches. The researchers discussed the main details of some significant studies and highlighted future research directions in Hadith field, including new sentiment-based applications.

As it can be noticed, many researches proposed various approaches and utilized different techniques to identify narrators' names. However, to the best of the author's knowledge, there is no research work process Isnads of Hadith using Genetic Algorithms (GA). In this paper, a novel approach for Isnad processing using GA is presented. This approach, which called "Isnad-GA", includes tagging of Isnad-Phrases according to predefined POIs, which are: Prophet Name, Narrator Name, Narrator-Name Prefix, Received Method, Received-Method Prefix, Replacement, Title and Others.

III. THE PROPOSED ISNAD-GA APPROACH

In this section, the proposed approach, which called "Isnad-GA", will be discussed in detail.

A. THE ARCHITECTURE OF THE PROPOSED APPROACH

Fig. 1 shows the architecture of the proposed approach, which based on GA. First, Hadiths have been collected from "Sahih Muslim" book [3] that considered as one of the most authentic and reputed resources for Hadiths; it is available online on many websites, such as [7], [8]. "Sahih Muslim" contains 3033 Hadiths without repetition (as many Hadiths may have more than one narration chain for the same Matn). All these Hadiths have been considered as the raw Hadiths that handled by Hadith scholar (expert) in the preprocessing phase, which will be discussed later. After that, GA will be used to manipulate Hadiths and predict the POIs for Isnad-Phrases.

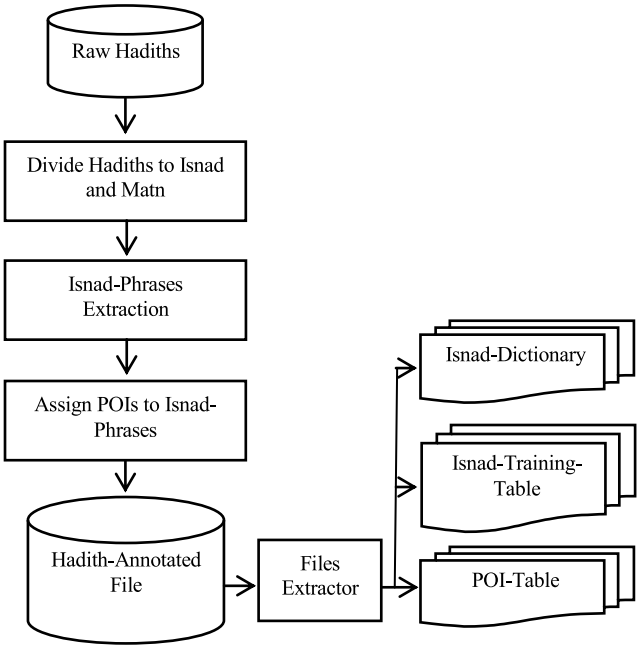

FIGURE 2. Preprocessing phase.

At the end, Isnads, Isnad-Phrases, and its corresponding POIs will be collected in one file called RecognizedPOI.

Fig. 2 shows the preprocessing phase, which starts with the division of Hadith into Isnad and Matn manually by Hadith scholar (expert). Then Isnads are arranged and stored in one file called All-Isnads-table; this step is also done manually by the expert. After that, Isnads divided randomly into 2730 Isnads to represent the training dataset (approximately nine-tenths of Isnads), and the rest of Isnads (303 Isnads) considered as testing dataset. The size of the training dataset has been chosen large enough to return statistically meaningful results and to represent the dataset as a whole, so there will be no different characteristics between training dataset and testing dataset. For training dataset, the expert extracted Isnad-Phrases from all Isnads and tagged them to the proper POIs; then, all Isnads, Isnad-Phrases, and corresponding POIs are stored in one file called Hadith-Annotated file. This file used to create three ''Aid-Files'' to Isnad which are:

1) Isnad-Dictionary, which contains:
   a) All distinct Isnad-Phrases of all Isnads in the training dataset.
   b) The corresponding POIs for these Isnad-Phrases.
   c) The number of frequencies for each pair of Isnad-Phrase and its POI.
2) Isnad-Training-Table, which contains all Isnads in the training dataset and the correct POIs sequence for each Isnad, which determines by Hadith scholar (expert).
3) POI-Table, which contains all distinct POIs of all Isnads in the training dataset and its frequencies.

With this technique, there is no need to do training every time a new data come; hence, a lot of time will be saved. These files will be used by GA to manipulate Hadiths and predict the POIs for all Isnads, as will be discussed later.

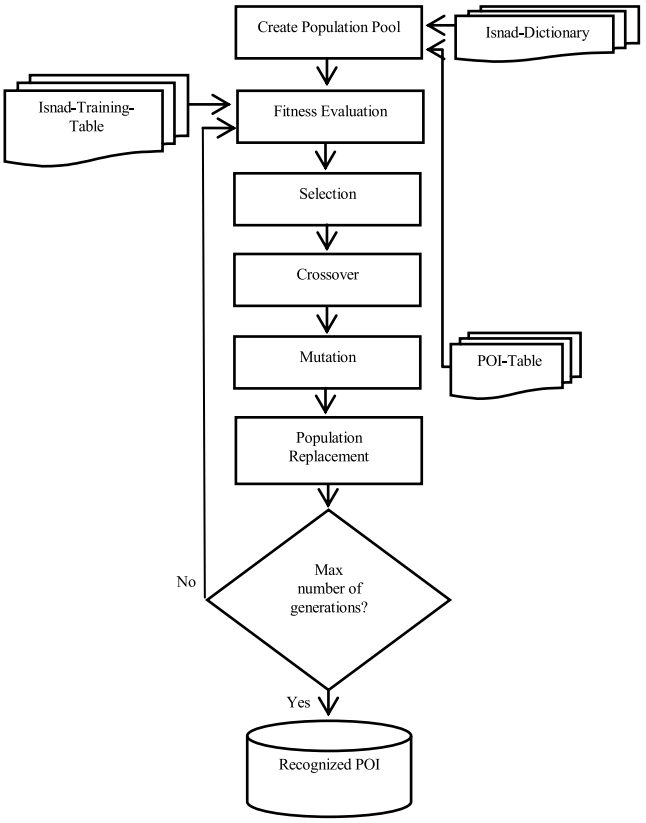

FIGURE 3. Genetic algorithm for POIs prediction.

B. REPRESENTATION OF GENETIC ALGORITHM

Fig. 3 depicts the usage of GA in the proposed approach (Isnad-GA); it shows the role of the three Aid-Files and the various operations of GA like selection, crossover and mutation.

1) CHROMOSOME REPRESENTATION

The Isnads-processing problem was represented in a form that can be implemented in GA; therefore, the proposed approach assumes that the ''Chromosome'' is the POIs sequence for one Isnad, and the ''Gene'' is one POI in this sequence, as represented in Table 4. For a certain Isnad, the Isnad-Dictionary is used to get all possible POIs for all Isnad-Phrases in this Isnad, these Isnad-Phrases, and its corresponding POIs are collected in one file called ''Possible-POI'' file.

2) CREATE POPULATION POOL

For a certain Isnad, the Possible-POI file used to generate chromosomes as follows:

- If the first Isnad-Phrase in the Isnad has only one possible POI in the Possible-POI file, then this POI will be considered as the first gene in the first chromosome.

TABLE 4. Chromosome representation for Hadith1.

Isnad	حَدَّثَنَا عُبَيْدُ اللهِ بْنُ سَعِيدٍ، وَعَبْدُ بْنُ حُمَيْدٍ، قَالاَ: حَدَّثَنَا أَبُو عَامِرٍ الْعَقَدِيُّ، حَدَّثَنَا سُلَيْمَانُ بْنُ بِلاَلٍ، عَنْ عَبْدِ اللهِ بْنِ دِينَارٍ، عَنْ أَبِي صَالِحٍ، عَنْ أَبِي هُرَيْرَةَ، عَنِ النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، قَالَ: Obaidullah bin Saeed narrated, and Abdo bin Humid, they said: Abu Amer Alaqdy narrated, Suleiman bin Bilal narrated, from Abdullah Ibn Dinar, from Abu Salih, from Abu Huraira that the Prophet peace be upon him said:					
Isnad-Phrases	حَدَّثَنَا (narrated)	عُبَيْدُ اللهِ بْنُ سَعِيدٍ (Obaidullah bin Saeed)	عَبْدُ بْنُ حُمَيْدٍ (Abdo bin Humid)	قَالاَ (they said)	حَدَّثَنَا (narrated)	…
Chromosome	Received Method	Narrator Name	Narrator Name	Received-Method Prefix	Received Method	…

- If the first Isnad-Phrase in the Isnad has more than one POI in the Possible-POI file, then the proposed approach will choose a random POI from the Possible-POI file and consider it as the first gene in the first chromosome.
- If the first Isnad-Phrase in the Isnad does not exist in the Possible-POI file, since it does not exist in the training dataset, then the proposed approach uses the Roulette Wheel algorithm to choose one POI from POI-Table. This POI will be considered as the first gene in the first chromosome.

This process will be repeated to create the remaining genes for the first chromosome. After the creation of all chromosomes in one population for a specific Isnad, these chromosomes are stored in a table called ''Population-Table'', as shown in Table 5.

The proposed method compares the multi-word narrator's name with the Isnad-Phrases in the ''Possible-POI'' file. If the narrator's name has only one possible POI in the Possible-POI file, then this POI will be assigned to this name. However, if the name has more than one POI in the Possible-POI file, then a random POI will be chosen from the Possible-POI file and assigned to this name. Finally, if the name does not exist in the Possible-POI file, since it does not exist in the training dataset, then the proposed approach uses the Roulette Wheel algorithm to choose one POI from POI-Table.

3) FITNESS FUNCTION

To evaluate the fitness of one chromosome in a specific population, the proposed approach uses the following equation:

$$Fitness(chromosome\ (x)) = \frac{occurrences\ (chromosome\ (x))}{\sum_{y=1}^{pop-size} occurrences\ (chromosome\ (y))} \quad (1)$$

where *chromosome(x)* is the chromosome whose fitness will be calculated, *occurrences (chromosome(x))* is the number of occurrences of the chromosome(x) in the Isnad-Training-Table, and *pop-size* is the size of the population.

TABLE 5. Example of population table for Hadith1's Isnad.

Hadith	حَدَّثَنَا عُبَيْدُ اللَّهِ بْنُ سَعِيدٍ، وَعَبْدُ بْنُ حُمَيْدٍ، قَالاَ حَدَّثَنَا أَبُو عَامِرٍ الْعَقَدِيُّ، حَدَّثَنَا سُلَيْمَانُ بْنُ بِلاَلٍ، عَنْ عَبْدِ اللَّهِ بْنِ دِينَارٍ، عَنْ أَبِي صَالِحٍ، عَنْ أَبِي هُرَيْرَةَ، عَنِ النَّبِيِّ صلى الله عليه وسلم قَالَ: "الإِيمَانُ بِضْعٌ وَسَبْعُونَ شُعْبَةً وَالْحَيَاءُ شُعْبَةٌ مِنَ الإِيمَانِ". Obaidullah bin Saeed narrated, and Abdo bin Humid, they said: Abu Amer Alaqdy narrated, Suleiman bin Bilal narrated, from Abdullah Ibn Dinar, from Abu Salih, from Abu Huraira that the Prophet peace be upon him said: "Iman (faith) has over seventy branches, and modesty is a branch of Iman".					
Isnad	حَدَّثَنَا عُبَيْدُ اللهِ بْنُ سَعِيدٍ، وَعَبْدُ بْنُ حُمَيْدٍ، قَالاَ: حَدَّثَنَا أَبُو عَامِرٍ الْعَقَدِيُّ، حَدَّثَنَا سُلَيْمَانُ بْنُ بِلاَلٍ، عَنْ عَبْدِ اللهِ بْنِ دِينَارٍ، عَنْ أَبِي صَالِحٍ، عَنْ أَبِي هُرَيْرَةَ، عَنِ النَّبِيِّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، قَالَ: Obaidullah bin Saeed narrated, and Abdo bin Humid, they said :Abu Amer Alaqdy narrated, Suleiman bin Bilal narrated, from Abdullah Ibn Dinar, from Abu Salih, from Abu Huraira that the Prophet peace be upon him said:					
Isnad-Phrases	حَدَّثَنَا (narrated)	عُبَيْدُ اللهِ بْنُ سَعِيدٍ (Obaidullah bin Saeed)	عَبْدُ بْنُ حُمَيْدٍ (Abdo bin Humid)	قَالاَ (they said)	حَدَّثَنَا (narrated)	…
Chromosome1	Received Method	Narrator Name	Narrator Name	Received-Method Prefix	Received Method	…
Chromosome2	Received-Method Prefix	Narrator Name	Title	Received Method	Received -Method Prefix	…
Chromosome3	Received-Method Prefix	Narrator Name	Narrator Name	Received Method	Received Method	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮

4) SELECTION

This process selects the chromosome according to its fitness value; it is based on the Roulette Wheel algorithm [40]. Every chromosome reserves space on the wheel depends on its fitness value. So if the fitness value of the first chromosome is 85% and the fitness value of the second chromosome is 75%, then the first chromosome will get more space in the wheel; thus, it will have a high probability to be selected.

5) CROSSOVER

The crossover process aims to generate new chromosomes for the next generation. The proposed approach used the one-point crossover method; this method selects a random crossover point, then the POIs that precede this point for both parents are exchanged to produce two different children.

6) MUTATION

This process aims to change one of the POI in a chromosome by another POI. It selects the new POI from the POI-Table, which contains all distinct POIs of all Isnads in the training dataset and its frequencies. This process depends on the mutation rate, which is often getting a small value such as 1%. The mutation process guides the algorithm in a specific direction in the search space; hence, it avoids the local solution and gives diversity in solutions to reach the global solution.

Algorithm 1 GAForPOI Function

```
1   Function Name: GAforPOI
2       Function Input: Mutation_Rate, Max_No_Of_Genarations, Population_Size, Isnad-Dictionary,
            Isnad-Training-Table, POI-Table, Testing_DataSet
3       Function Output: RecognizedPOI
4       START
5       RecognizedPOI = NULL
6           FOR x = 1 to Number of Isnads in Testing_DataSet
7           Possible-POI = GetPossiblePOI(Isnad, Isnad-Dictionary)
8           Population_Table = CreatePopulation(Possible-POI, Population_Size )
9               FOR Generation = 1 to Max_No_Of_Genarations
10              GenerationFitness = FitnessFunction(Population_Table)
11                  FOR y = 1 to half of Population_Size
12                  FirstSelectedChromosome = SelectedByRoulette Wheel(Population_Table)
13                  SecondSelectedChromosome = SelectedByRoulette Wheel(Population_Table)
14                  New_Chromosome1,New_Chromosome2← DoCrossover(FirstSelectedChromosome,
15                  SecondSelectedChromosome)
16                  Child1 = DoMutation(New_Chromosome1, Mutation_Rate)
17                  Child2 = DoMutation(New_Chromosome2, Mutation_Rate)
18                  END FOR
19              CallFunction: IsnadPopulationReplacment
20              END FOR
21          GenerationFitness = FitnessFunction(Population_Table)
22          Consider POIs of the fittest chromosome as POIs for Isnad
23          Add Isnad, its Isnad-Phrases and the fittest chromosome to RecognizedPOI.
24          END FOR
25      Return RecognizedPOI
26      END
```

C. IMPLEMENTATION

Algorithm 1 explains the ''GAForPOI'' function that implements the main operators of GA; line 2 shows the inputs of the algorithm, which are: Mutation_Rate, Max_No_Of_Genarations, Population_Size, Isnad-Dictionary, Isnad-Training-Table, POI-Table and Testing_DataSet, whereas the output is the RecognizedPOI file as shown in line 4. The algorithm supposes that the Recognized-POI file initially contains ''Null'' value, as shown in line 6. Line 7 shows the FOR statement that aims to go through all Isnads in the testing dataset and then predicts the POIs for all Isnad-Phrases in each Isnad. At the end, the Isnads, Isnad-Phrases and corresponding POIs will be added to the RecognizedPOI file.

Line 8 depicts the creation process of the Possible-POI table, which contains a list of all possible POIs for each Isnad-Phrase in a specific Isnad. This process checks the Isnad-Dictionary that contains all the Isnad-Phrases in the training dataset and their corresponding POIs. This creation process involves calling the ''GetPossiblePOI'' function that accepts Isnad and Isnad-Dictionary as inputs and returns Possible-POI table as output. Line 9 depicts the population creation process for a certain Isnad. As explained before, this process includes calling the ''CreatePopulation'' function that accepts Possible-POI, POI-Table and Population_Size as inputs, and return Population-Table as output. The Population-Table contains all chromosomes that represent the population for one Isnad.

The algorithm after that creates many generations according to the Max_No_of_Generations value, as explained from line 10 to line 21. For each generation, the algorithm first calculates the fitness of all chromosomes in this generation using Eq. (1), as shown in line 11, then it applies the select, crossover and mutation processes for this population to create a new generation. This process implies a loop for half the population size, as explained from line 12 to line 19. The old population will be replaced with the new one, as shown in line 20. If the algorithm does not reach the Max_No_Of_Generations, it will go back to the fitness evaluation step (line 11) and repeats all previous steps to create a new generation. After the algorithm reaches the Max_No_Of_Generations, it will evaluate the fitness of the last generation, as shown in line 22. The algorithm will consider the sequence of POIs of the fittest chromosome as POIs for Isnad-Phrases of the current Isnad. Then it will add the current Isnad, its Isnad-Phrases, and the fittest chromosome to the RecognizedPOI file.

Finally, the algorithm will repeat all the previous steps for each Isnad in the testing dataset, and then it will generate the Recognized-POI file as the output file of the algorithm, as shown in line 26. The Recognized-POI file will be used to calculate the accuracy of the proposed approach; this file

TABLE 6. Experimental parameters.

Parameter	Values of parameter
Corpus size	3033 Hadiths
Training dataset	2730 Hadiths
Testing dataset	303 Hadiths
Selection method	Roulette Wheel method
Crossover method	One-point Crossover
Population size	20, 40, 60, 80
Maximum No. of Generations	5, 10, 20
Mutation rate	1%, 5%, 10%

will be compared with a new copy of the testing dataset created by tagging all its Isnad-Phrases correctly by the expert.

IV. THE EXPERIMENTAL RESULTS

As explained earlier, the corpus was considered as the whole ''Sahih Muslim'' book. This book deemed as one of the most authentic and reputed resources for Hadiths. The researcher collected all Hadiths narrated in this book, which are 3033 Hadiths without repetition (as many Hadiths may have more than one narration chain for the same Matn). The book is available online on many websites, such as [7], [8]. Hadith scholar (expert) had divided Hadiths manually into Isnads and Matns; after that, Isnads were divided randomly into 2730 Isnads to represent the training dataset (approximately nine-tenths of Isnads), and the rest of Isnads (303 Isnads) considered as testing dataset.

All Isnad-Phrases of the Isnads in the training dataset were tagged to the proper POI by the expert; the testing dataset remained without tagging. Another copy of the testing dataset was created by tagging all its Isnad-Phrases correctly, this new (correct) testing dataset used by the proposed approach to calculate the accuracy; the approach predicts the POIs for the original testing dataset and compares the results with the correct testing dataset.

The parameters used in the experiments of this research are shown in Table 6; the corpus size is equal to 3033 Hadiths, the training dataset contains 2730 Hadiths, and the testing dataset contains 303 Hadiths. The selection method used in this research is the Roulette Wheel method, and the crossover method is the one-point crossover method. The population sizes are 20, 40, 60, and 80, the maximum numbers of generations are 5, 10, and 20, and the mutation rates are 1%, 5%, and 10%.

Table 7 shows the experimental results; the first column contains the values that considered as population size, while the second column contains the values of the ''maximum number of generations'', and the third column contains the values of the mutation rate. Therefore the experiments conducted on all possible combinations of these values. The last column shows the accuracy of the proposed ''Isnad-GA'' approach. It is noticed that if the population size is 20, then the best accuracy will be 77.87%; this result is attained when the ''maximum number of generations'' is 20, and the mutation rate is 1%.

TABLE 7. Experimental results.

Population Size	Maximum number of generations	Mutation Rate	Accuracy
20	5	1%	77.52%
		5%	76.28%
		10%	75.73%
	10	1%	77.81%
		5%	77.32%
		10%	75.97%
	20	1%	**77.87%**
		5%	77.30%
		10%	76.11%
40	5	1%	78.27%
		5%	77.33%
		10%	76.29%
	10	1%	79.05%
		5%	78.72%
		10%	77.56%
	20	1%	**81.44%**
		5%	81.15%
		10%	79.06%
60	5	1%	78.18%
		5%	77. 23%
		10%	77.15%
	10	1%	78.80%
		5%	78.23%
		10%	78.75%
	20	1%	**81.16%**
		5%	80.10%
		10%	79.52%
80	5	1%	78.02%
		5%	77.11%
		10%	76.85%
	10	1%	**81.11 %**
		5%	80.27 %
		10%	79.64%
	20	1%	80.08 %
		5%	79.87%
		10%	77.33 %

If the population size is 40, then the best accuracy will be 81.44%; this result is attained when the ''maximum number of generations'' is 20, and the mutation rate is 1%. If the population size is 60, then the best accuracy will be 81.16%; this result is attained when the ''maximum number of generations'' is 20, and the mutation rate is 1%. If the population size is 80, then the best accuracy will be 81.11%; this result is attained when the ''maximum number of generations'' is 10, and the mutation rate is 1%.

Consequently, it is concluded that the ''best'' accuracy which the proposed approach achieved was 81.44%. It is observed that, in most cases, if the ''maximum number of

generations'' increased, the accuracy would increase too. Finally, if the mutation rate is 1%, the proposed approach will give the best results in most cases; in contrast, the proposed approach will present the worst results if the mutation rate is 10%.

V. CONCLUSION

A novel approach called ''Isnad-GA'' was introduced in this paper for processing Isnad of Hadith using Genetic Algorithms (GA). The proposed approach considered all Hadiths narrated in ''Sahih Muslim'' book, which are 3033 Hadiths without repetition, as the corpus of this research study. Isnads of Hadiths have been divided randomly into 2730 Isnads to represent the training dataset, and 303 Isnads to represent the testing dataset. This approach proved that GA could be used to predict the narrators' names and the other POIs for Isnad.

The main factors that affect the results of the proposed approach, like population size, maximum number of generations, and mutation rate, have been discussed in detail. The results showed that if the population size is 40, the maximum number of generations is 20, and the mutation rate is 1%; then, the proposed approach will achieve the best accuracy, which is 81.44%. These results are not as high as expected; therefore, in future work, one of the aims will be enhancing the ''Isnad-GA'' method by examining several selection and crossover techniques, such as the one-child selection technique, and the two-point crossover technique. Another direction could be the investigation of utilizing other approaches for Isnad processing, such as deep learning.

ACKNOWLEDGMENT

The author greatly thanks Dr. Naser Younes Sabra, Ph.D. in Islamic Studies and Hadith sciences, for his great help and support, without him, this research would not have completed. As well, the author would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] M. Al-Tahhan and T. M. Al Hadith, *Clarification of Hadith's Terminology*. Riyadh, United Arab Emirates: Maarif Library for Publishing and Distribution, (in Arabic), 2011.
[2] M. M. Al-Azami, *Studies in Hadith Methodology and Literature*. Indianapolis, IN, USA: American Trust, 1978.
[3] M. Al-Hajjaj, *Sahih Muslim*. Cairo, Egypt: Dar Ibn Al-Jawzi Publication and Distributors, (in Arabic), 2009.
[4] A. Diaz-Manríquez, A. B. Ríos-Alvarado, J. H. Barrón-Zambrano, T. Guerrero-Melendez, and J. Elizondo-Leal, ''An automatic document classifier system based on genetic algorithm and taxonomy,'' *IEEE Access*, vol. 6, pp. 21552–21559, Mar. 2018.
[5] H.-P. Ren, X.-N. Huang, and J.-X. Hao, ''Finding robust adaptation gene regulatory networks using multi-objective genetic algorithm,'' *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 3, pp. 571–577, May 2016.
[6] M. Sepahvand, F. Abdali-Mohammadi, and F. Mardukhi, ''Evolutionary metric-learning-based recognition algorithm for online isolated persian/arabic characters, reconstructed using inertial pen signals,'' *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2872–2884, Sep. 2017.
[7] *Islam Web*. Accessed: Jul. 15, 2018. [Online]. Available: http://www.islamweb.net
[8] *Shamela*. Accessed: Aug. 25, 2018. [Online]. Available: http://shamela.ws
[9] *Dorar*. Accessed: Aug. 28, 2018. [Online]. Available: http://www.dorar.net
[10] F. Harrag and A. Hamdi-Cherif, ''UML modeling of text mining in Arabic language and application to the prophetic traditions, hadiths,'' in *Proc. 1st Int. Symp. Comput. Arabic Lang. Exhib. (ISCAL)*. Riyadh: Saudi Arabia: KSA, 2007, pp. 11–20.
[11] F. Harrag, A. Hamdi-Cherif, and E. El-Qawasmeh, ''Information retrieval architecture for hadith text mining,'' *J. Digit. Inf. Manage.*, vol. 6, no. 6, pp. 449–455, 2008.
[12] F. Harrag, A. Hamdi-Cherif, and E. El-Qawasmeh, ''Vector space model for Arabic information retrieval—Application to 'Hadith' indexing,'' in *Proc. 1st Int. conf. Appl. Digit. Inf. Web Technol. (ICADIWT)*, Ostrava, Czech Republic, 2008, pp. 107–112.
[13] F. Harrag and E. El-Qawasmah, ''Neural network for arabic text classification,'' in *Proc. 2nd Int. Conf. Appl. Digit. Inf. Web Technol. (ICADIWT)*, Luton, U.K., Aug. 2009, pp. 778–783.
[14] F. Harrag, A. Alothaim, A. Abanmy, F. Alomaigan, and S. Alsalehi, ''Ontology extraction approach for prophetic narration (hadith) using association rules,'' *Int. J. Islamic Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 48–57, 2013.
[15] F. Harrag, ''Text mining approach for knowledge extraction in Sahîh Al-Bukhari,'' *Comput. Hum. Behav.*, vol. 30, pp. 558–566, Jan. 2014.
[16] K. Faidi, R. Ayed, I. Bounhas, and B. Elayeb, ''Comparing Arabic NLP tools for Hadith classification,'' *Int. J. Islamic Appl. Comput. Sci. Technol.*, vol. 3, no. 3, pp. 1–12, 2015.
[17] M. A. Saloot, N. Idris, R. Mahmud, S. Ja'Afar, D. Thorleuchter, and A. Gani, ''Hadith data mining and classification: A comparative analysis,'' *Artif. Intell. Rev.*, vol. 46, no. 1, pp. 113–128, Jun. 2016.
[18] F. Harrag, E. El-Qawasmah, and A. M. S. Al-Salman, ''Stemming as a feature reduction technique for Arabic text categorization,'' in *Proc. 10th Int. Symp. Program. Syst. (ISPS)*, Ghardaia, Algeria, Apr. 2011, pp. 128–133.
[19] A. Mahmood, H. U. Khan, Zahoor-Ur-Rehman, and W. Khan, ''Query based information retrieval and knowledge extraction using hadith datasets,'' in *Proc. 13th Int. Conf. Emerg. Technol. (ICET)*, Dec. 2017, pp. 1–6.
[20] M. Ghazizadeh, M. H. Zahedi, M. Kahani, and B. Minaei Bidgoli, ''Fuzzy expert system in determining Hadith[1] validity,'' in *Advances in Computer and Information Sciences and Engineering*, T. Sobh, Eds. Dordrecht, The Netherlands: Springer, 2008, pp. 354–359.
[21] A. Azmi and N. Bin Badia, ''ITree–automating the construction of the narration tree of Hadiths (prophetic traditions),'' in *Proc. 6th Int. Conf. Natural Lang. Process. Knowl. Eng. (NLPKE)*, Beijing, China, Aug. 2010, pp. 1–7.
[22] A. Azmi and N. AlBadia, ''Mining and visualizing the narration tree of Hadiths (prophetic traditions),'' in *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches*. Hershey, PA, USA: IGI Global, 2012, pp. 239–257, ch. 16.
[23] A. Azmi and N. B. Badia, ''e-Narrator—An application for creating an ontology of Hadiths narration tree semantically and graphically,'' *Arabian J. Sci. Eng.*, vol. 35, no. 2C, pp. 51–68, 2010.
[24] I. Bounhas, B. Elayeb, F. Evrard, and Y. Slimani, ''Toward a computer study of the reliability of Arabic stories,'' *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 8, pp. 1686–1705, 2010.
[25] I. Bounhas, B. Elayeb, F. Evrard, and Y. Slimani, ''Information reliability evaluation: From Arabic storytelling to computer sciences,'' *ACM J. Comput. Cultural Heritage*, vol. 8, no. 3, 2015, Art. no. 14.
[26] K. Aldhaln, A. Zeki, A. Zeki, and H. Alreshidi, ''Improving knowledge extraction of Hadith classifier using decision tree algorithm,'' in *Proc. Int. Conf. Inf. Retrieval Knowl. Manage.*, Mar. 2012, pp. 148–152.
[27] A. Azmi and A. AlOfaidly, ''A novel method to automatically pass hukm on hadith,'' in *Proc. 5th Int. Conf. Arabic Lang. Process. (CITALA)*, Morocco, IN, USA, 2014, pp. 118–124.
[28] M. Siddiqui, M. Saleh, and A. Bagais, ''Extraction and visualization of the chain of narrators from hadiths using named entity recognition and classification,'' *Int. J. Comput. Linguistics Res.*, vol. 5, no. 1, pp. 14–25, 2014.
[29] M. Alhawarat, ''A domain-based approach to extract Arabic person names using n-grams and simple rules,'' *Asian J. Inf. Technol.*, vol. 14, no. 8, pp. 287–293, 2015.

[30] N. A. Rahman, N. Alias, N. K. Ismail, Z. B. M. Nor, and M. N. B. Alias, "An identification of authentic narrator's name features in Malay Hadith texts," in *Proc. IEEE Conf. Open Syst. (ICOS)*, Bandar Melaka, Malaysia, Aug. 2015, pp. 79–84.
[31] S. R. M, Najib, N. A, Rahman, N. K, Ismail, N. Alias, Z. M. Nor, and M. N. Alias, "Comparative study of machine learning approach on Malay translated Hadith text classification based on Sanad," in *Proc. MATEC Web Conf.*, Selangor, Malaysia, vol. 135, 2017, Art. no. 66.
[32] S. S. Balgasem and L. Q. Zakaria, "A hybrid method of rule-based approach and statistical measures for recognizing narrators name in hadith," in *Proc. 6th Int. Conf. Electr. Eng. Inform. (ICEEI)*, Langkawi, Malaysia, Nov. 2017, pp. 1–5.
[33] K. Bilal and S. Mohsin, "Muhadith: A cloud based distributed expert system for classification of Ahadith," in *Proc. 10th Int. Conf. Frontiers Inf. Technol.*, Dec. 2012, pp. 73–78.
[34] M. Najeeb, A. Abdelkader, and M. Al-Zghoul, "Arabic natural language processing laboratory serving Islamic sciences," *Int. J. Adv. Comput. Sci. Applic.*, vol. 5, no. 3, pp. 114–117, 2014.
[35] M. M. A. Najeeb, "XML database for Hadith and narrators," *Amer. J. Appl. Sci.*, vol. 13, no. 1, pp. 55–63, Jan. 2016.
[36] M. Najeeb, "Muti-agent system for Hadith processing," *Int. J. Softw. Eng. Appl.*, vol. 9, no. 9, pp. 153–166, 2015.
[37] A. Mahmood, H. U. Khan, F. K. Alarfaj, M. Ramzan, and M. Ilyas, "A multilingual datasets repository of the Hadith content," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 165–172, 2018.
[38] I. Bounhas, "On the usage of a classical arabic corpus as a language resource: Related research and key challenges," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 18, no. 3, pp. 1–45, Jan. 2019.
[39] 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, no. 2, pp. 1369–1414, Aug. 2019.
[40] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA, USA, Addison-Wesley, 1989.

MOATH MUSTAFA AHMAD NAJEEB received the B.Sc. degree in computer science from Al al-Bayt University, Jordan, in 2000, the M.Sc. degree in computer science from Amman Arab University, Jordan, in 2006, and the Ph.D. degree in computer information systems from the University of Banking and Financial Sciences, Jordan, in 2011. From 2000 to 2004, he worked as a Web Developer, Systems Analyst and Oracle Database Administrator in UAE. He worked as an Assistant Professor with the Computer College, Jadara University, Jordan, in 2011. He is currently an Assistant Professor with the Computer College, Al-Qunfudah, Umm Al-Qura University, Saudi Arabia. His research interests include Arabic natural language processing, pattern recognition, and data mining.