ISSN 2278-3091
Volume 8, No.1.5, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse5881.53019.pdf
https://doi.org/10.30534/ijatcse/2019/5881.52019

A Novel Graph-Based Representation for Hadith *Sanad*

Emha Taufiq Luthfi[1], Nanna Suryana[2], Abdulsamad Hasan Basari[3]
[1]Universitas AMIKOM Yogyakarta, Indonesia, emhataufiqluthfi@amikom.ac.id
[2]Universiti Teknikal Malaysia Melaka, Malaysia, nsuryana@utem.edu.my
[3]Universiti Teknikal Malaysia Melaka, Malaysia, abdsamad@utem.edu.my

ABSTRACT

Hadith is the secondary source of Islam legislation that has three-part, i.e., *Sanad*, *Matn*, and *Taraf*. *Sanad* is an essential part of Hadith that represent the chain of Narrator who conveys the Hadith. Based on Hadith Science, the authentication of the Hadith also could be observed through the state of its *Sanad*. Most studies on the Hadith *Sanad* representation apply an ontology and XML. Thus, this study proposed a new model of *Sanad* Hadith representation exploits the Graph model. First, the candidate of Graph node and Graph relation were extracted automatically from raw Arabic Hadith text using Arabic Part of Speech (A-POS) and Arabic Named Entity Recognition (A-NER). Then, a novel machine learning model for the Hadith *Sanad* Graph Construction developed employs SVM and GBM algorithm. That model attained the best performance on 0.84 and 0.92 precision average, 0.83 and 0.91 recall average, 0.82 and 0.90 f1-score average. The final result of this study was a Hadith *Sanad* Graph that had been verified the correctness compare with the original Hadith text.

Keywords: Knowledge Representation, Knowledge Graph, Hadith, Hadith Representation, Hadith *Sanad* Representation, Hadith Graph, Hadith *Sanad* Graph.

1. INTRODUCTION

Today's the world was growing very fast in step with the growth of the internet and online media. Nowadays most people including Muslims used online media as a primary source of information or knowledge. Most Muslims adopted online media as the primary reference for exploring religious content including when seeking for verse of Qurán or Hadith. The problem is not all information or knowledge on the online media was verified its correctness. There are also spreading wide on online media an Islamic content that not verified its accuracy [1]. Al-Quran and Hadith are significant resources of Islamic law which every Muslims all around the world must refer [2][3]. Quran is the most authentic and unaltered Holy book of God over more than 14 centuries ever since revelated [4]. Hadith is Islamic law that originated from the collection of saying, action, decision or characteristic of Prophet Muhammad PBUH [5][6]. Unlike the Holy Quran, Hadith that spread amid Muslims are not all authentic [7]. Thus, Muslims needs to authenticate the correctness of Hadith, especially when accessed from online media

Hadith science (عِلْمُ الحَدِيْثِ) is one of Islamic learning that utilized to inspect and characterize the validity of Hadith [8]. According to Hadith science, every Hadith has three-part [7][9] as shown in Figure 1, viz:

1. *Sanad* (اَلسَّنَدُ)
   Sanad is the sequence of narrators (اَلرِّجَالُ) that direct to the Hadith text. The *Sanad* contains all those who narrated the version, start from the last narrator and ending on the Prophet (PBUH) [8] [10][11][12].
2. *Matn* (اَلْمَتْنُ)
   Text or content of Hadith delivered by *Sanad* [12][13].
3. *Taraf* (الطرف)
   The part, or sentence opener, on the *Matn* that indicate to the statement, deeds or characteristics of the Prophet (PBUH), and his agreement to others deed [14].

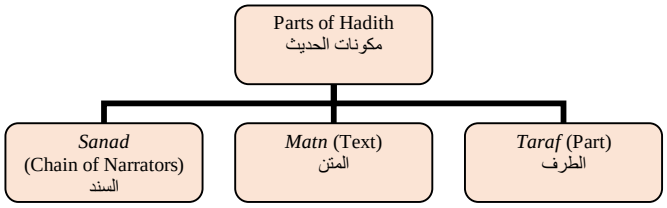


Figure 1: Parts of Hadith

Hadith can be accepted (*Maqbûl*) as correct when complying with several guidelines which are [15][16]:
Entire involved narrators in *Sanad* Hadith were qualified (*'Adil* and *Dhabith*).

1. Entire *Sanad* of Hadith continues.
2. *Matn* of Hadith was free of *syadz*.
3. *Matn* of Hadith was free from *'illat*.

In case one criterion of the accepted Hadith (*Maqbûl*) is not fulfilled, the status of Hadith would change into rejected (*Mardûd*) or not correct.

This article would concentrate on explaining and expose the representation of *Sanad* Hadith as a foundation for Hadith authentication partly based on the criteria point 1 and 2. Although extensive academic research has explored the portrait of *Sanad* Hadith [8][17][18][19][20][21][22][23][24], less research has investigated the *Sanad* Hadith representation with the utilize of Graph model. The rest of the article is structured as follows: First, the previous literature on the *Sanad* Hadith representation including its critical analysis. This view is followed by a theoretical explanation of the knowledge graph and description of the research methodology used in the study. The results of this research inquiry are then discussed. Finally, implications, limitations, and directions for future research are explained.

2. PREVIOUS WORK

In recent years, experts have dedicated the effort work to the research on the Islamic knowledge representation. On the domain of Hadith, exist several studies on Hadith representation, but slightly research can be found for the Graph representation of the Hadith *Sanad*. The research conducted by [8] proposed a new lexicon model for Hadith representation by utilizing HPSG formalism. Each part of Hadith *Isnad* like Narrator, telling tool is represented in a separate XML document with HPSG lexical features targeted explicitly at morphological analysis. The HPSD grammar rule can help to define the form and the relation of each part of Hadith. The study that undertakes by [17] proposed a text representation graph for the chain of narrators in hadith texts. This study extracted 18 hadith texts and produced 82 narrator names and 85 relationships between the narrators on the graph. The result shows that all 18 hadith node in the graph is the same as the original. However, this study used only Malay Hadith text with the assumption that the Hadith text as an input has the same format to be extracted. And incapable of identifying distinct node names but the same Narrator. Incompetent to recognize identical node names but the different Narrator.

Research that carried out by [18] focuses on structuring Digital Hadith Text to describes its textual part using the Text Encoding Initiative (TEI) standard encoding. Each Narrator on *Isnad* encoded in <persName> element with the "xml:id" to kept the Narrator order. And *Matn* encoded in <p> element or <quote> element. Utilize 1000 Hadith text; the result shows the value of precision is 0.86. Also, the recall is 0.85 and an F-measure equal to 0.85. This study provides new insights into how to marking of most textual structures and explain the nature of the Digital Hadith text. However, the flexibility of the TEI schemes depend on the interpretation of strings embedded in the attributes of SGML tags created; These strings are not part of minimal SGML systems and need to defined correctly. The study performed by [19] proposes a new ontology model of the prophetic domain with the data collected from the Holy Quran, Al-Hadith and books correlated with the prophetic domain. As a result, this study outcome was the prophetic ontology model with 151 classes that organizing 1230 Arabic expressions, 210 object properties to relate individuals to individuals, 44 data property to relate individuals to literals and 825 individual words. This study limitation was that the object relations built in an entirely manual process using Arabic dictionaries.

The study that undertakes by [20] proposed a semantic model of all Islamic knowledge based on Holy Qurán, Hadith, *Ijma'*, and *Qiyas*. This study utilized an ontology implement with Protege to produce the semantic representation of Islamic knowledge (Quran, Hadith, *Ijma*, *Qiyas*). As a result, the ontology of Islamic Legislative Ontology was presented but still partially. The limitation of this study is the manual development of its ontology in consequence of the complication of the Arabic expressions. The study conducted by [21] proposed the Hadith Commentary Ontology. This study managed the Hadith explanation by a scholar, its relations to other parts of the Holy Scriptures and also variations of hadith that recited by a diverse Narrator with a distinct *Matn*. [22] proposed TibbOnto, a domain-dependent ontology on Al-Tibb Al-Nabawi. TibbOnto is built following five steps of a proofed ontology methodology, including utilizing the chapters Kitab Al-Tibb of Sahih Al- Bukhari as a source. As a result, TibbOnto built having thirteen classes and two subclasses of the domain.

Research that carried out by [23] proposed Multilingual Hadith Corpus (MHC) by utilizing Arabic, English, French, and Russian Hadith Text. The MHC built with a fully manual annotation process on the XML schema model. [25] proposed a domain-dependent ontology on zakat Hadith named the OntoHadith. The OntoHadith embody six modules which are: (1) Knowledgebase, (2) Inference Engine, (3) Ontological Dictionary, (4) Linguistics, (5) Matcher, and (6) SPARQL Query Engine. The OntoHadith model succeeded in assisting the retrieving hadiths of zakat on high accuracy (0.81 precision and 0.93 recall). The study conducted by [24] develops a representation of Hadith in the form of ontology as a model for the basis for Hadith *Isnad* judgment. The Narrator became the central concept since the narrator is the main constituent of Isnad. The ontology built captured all the properties and relationships of a narrator as indicated and detailed in the various books of Hadith. Table 1 points out details on each prior studies.

Table 1: Prior Hadith Representation Model

Citation	Input/Method	Strength	Weakness
[8]	**Input**: Arabic Hadith Text **Method**: Head-Driven Phrase Structure Grammars (HPSG) formalism **Output**: XML Documents of Hadith Lexicon	The HPSD grammar rule can help to define the form and the relation of each part of Hadith.	Needed tremendous effort to extract Hadith parts manually
[17]	**Input**: 18 Malay Text Hadith **Method**: Graph **Output**: The graph Model of Hadith chain of Narrator	100% accuracy for the Hadith node	Malay Hadith Text Manual pre-process on Hadith Text Incapable of identifying distinct node names but the same Narrator Incompetent to recognize identical node names but a different person of Narrator.
[18]	**Input**: Digital Arabic Hadith Text **Method**: Text Encoding Initiative (TEI) **Output**: Digital Arabic Hadith Text formatted with TEI standard encoding.	TEI which is XML-based allows the flexible marking of most textual structures and show the nature of the text.	The TEI schemes depend on the interpretation of strings embedded in the attributes of SGML tags created; These strings are not part of minimal SGML systems and need to defined correctly.
[19]	**Input**: The Holy Quran, Al-Hadith and books correlated to the prophetic domain. **Method**: Ontology **Output**: An ontology model of the prophetic domain.	The prophetic ontology is built having a complete glossary of terms (concepts, instances, and properties) with its explanations.	All object relations was built manually using Arabic dictionaries.
[20]	**Input**: Arabic Hadith Text from Shahih Bukhari Book volume I **Method**: Ontology **Output**: Islamic legislative Ontology	If completely done, enables the indirect linkages between four Islamic knowledge sources that Quranic, Hadith, *Ijma'* and *Qiyas*.	The ontology model was built fully manual due to the complexity of the Arabic language.
[21]	**Input**: Hadith Commentary, Quranic Verse **Method**: Ontology **Output**: Hadith Commentary Ontology	The ontology built enables the indirect linkages of Hadith and Quranic verses.	All object and its relations were built manually.
[22]	**Input**: An authentic Tibb Al-Nabawi Hadith text **Method**: Ontology **Output**: The TibbOntoModel (A domain-dependent ontology on Al-Tibb Al-Nabawi)	The TibbOnto is built with a standard methodology that proofed.	The scope was limited on the Al-Tibb Al-Nabawi domain, possible to extend.
[23]	**Input**: Arabic, English, French and Russian Hadith Text **Method**: An XML schema with annotation **Output**: Multilingual Hadith Corpus (MHC) on XML Schema format	Utilize four comprehensive languages of Hadith Text (Arabic, English, French, and Russian)	The corpus (MHC) built manually.
[25]	**Input**: The Hadiths of zakat **Method**: Ontology **Output**: The OntoHadith Model (A domain-dependent ontology on Zakat)	The OntoHadith model succeeded in assisting the retrieving hadiths of zakat on high accuracy (0.81 precision and 0.93 recall).	The scope was limited on the zakat domain, possible to extend.
[24]	**Input**: Hadith book of Ibn Hajar **Method**: Ontology **Output**: Hadith Isnad Ontology	Hadith *Isnad* Ontology built has a comprehensive property to support the knowledge needed to judge *Isnad*	A domain ontology does not represent a broad conceptual model, except only a model that is authentic for a particular domain.

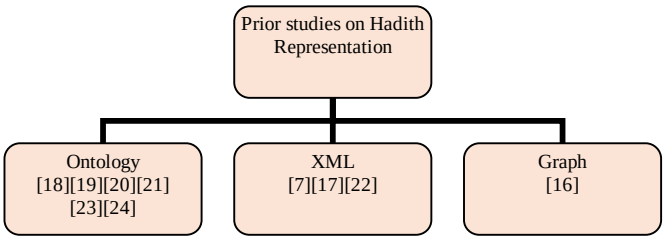


Figure 2: Method used on Prior Research

As can be observed from Table 1 and Figure 2 (above), the method employed on prior studies of Hadith representation gathers below three ways that are Ontology, XML and Graph. Graph model representation already utilized by [17]; however, that study was limited in several ways as detail in Table 1.

3. KNOWLEDGE GRAPH

One study by [26], describes that the promising future approach for Knowledge Representation (KR) is Informledge System (ILS) although the current grew up model is the Knowledge Graph (KG) include with Graph Database as shown on the evolution of KR techniques in Figure 3. This view is reinforced by [27] as displayed in Figure 4. Most of the big social networks like Google, Facebook, and Twitter implement KG included graph databases. The graph database is any repository model that employs graph composition with nodes and edges, to declare and save data [28]. Its power is on the boost of performance, flexibility, and agility when the data is structured on the connected graph. A graph database is one best choice to handle the complex, semi-structured, and densely connected data with very fast in terms of queries and gives a response in milliseconds [27].

A (labeled) property graph form is the common mostly adopted form of graphs in the context of graph databases [29]. A property graph builds up with a component of nodes, relationships, and properties. The graph database modeling possibly follows some designs pattern among is [30]:

1. Linked List
2. Multiple Relationships
3. Tags and Categories
4. Multi-Level Tree
5. R-Tree (spatial)
6. Activity Stream
7. Anti-pattern: Unconnected graph

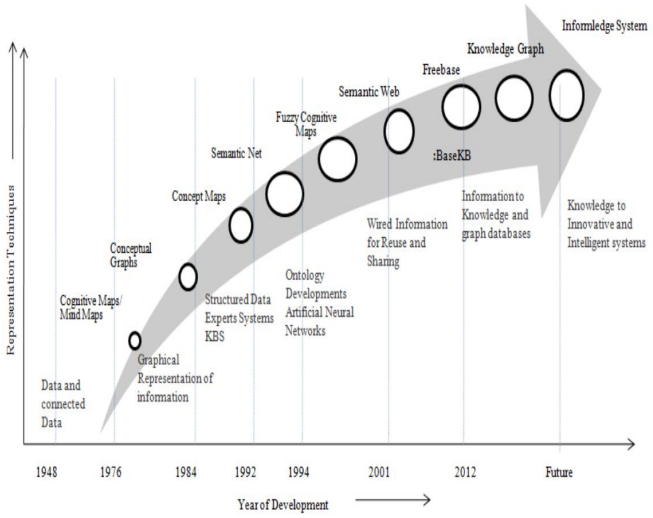


Figure 3: Evolution of Knowledge Representation Techniques [26]

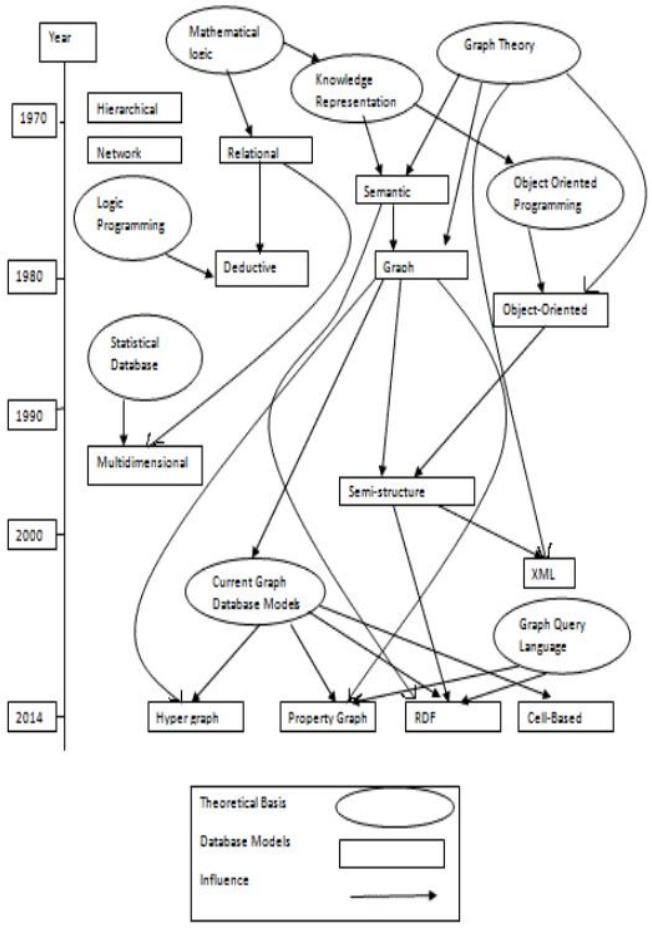


Figure 4: Evolution of Database Model [27]

4. PROPOSED MODEL

4.1 Methodology

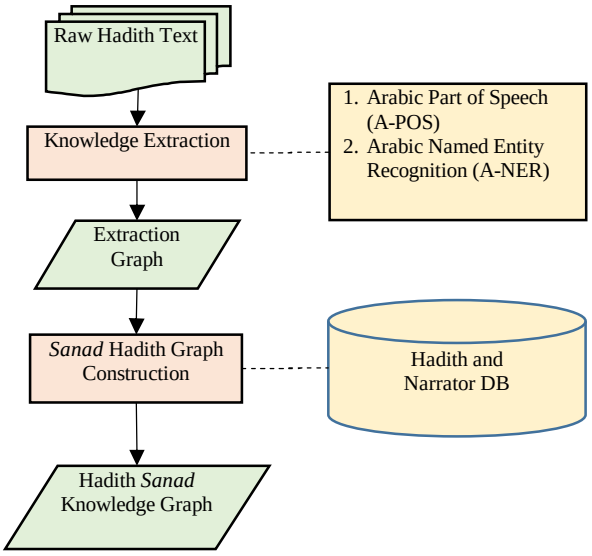


Figure 5: Methodology

The methodological approach taken in this study is a mixed methodology based on the stated strategy of [31] summarized in Figure 5. The first step is to gather and prepare raw Arabic Hadith text as an input for the model. Totally thirty of raw Hadith text from Shahih Bukhari, Shahih Muslim, and Jami' al-Tirmidzi collected from the Islamic web application http://qaalarasulallah.com/. One strength of this Islamic web application is provided authentic Arabic Hadith text (*Matn*), translation (English and other languages) and interactive chain of narrators (*Sanad*) for all main hadith compilation [32]. The specimen of raw Hadith text shown in Figure 6.

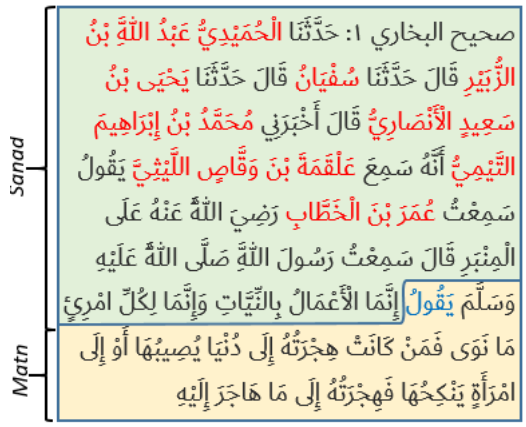


Figure 6: Example of Raw Arabic Hadith Text

The second step is extracting the knowledge from raw Arabic Hadith text intended to capture the candidate of Node and Relation for Graph. The Arabic Part of Speech (A-POS) and Arabic Named Entity Recognition (A-NER) utilized in this process. The extraction process has two principles [31]. First, if the extracted word is a noun, it will become candidate Graph Nodes. Second, if the retrieved word is a verb, it will become candidates for Graph Relations. This article utilized the A-POS and A-NER adopted from [33] including the Tag-Set. A Noun as Graph Nodes candidate would be taken from the tag B-PER and I-PER that show شخص (Person - Noun). A Verb as Graph Relation candidates would be considered from the tag PSTV or PSTV+PRO that indicate فعل ماضي or فعل ماضي + ضمير (Past Tense Verb + Person Pronouns). The key strengths of [33][34][35][36][37] were proposing the NLP model of word segmentation, A-POS, and A-NER as a single processing element with high performances result. Figure 7 and Figure 8 show the specimen output of A-POS and A-NER process.

صحيح_B-MISC البخاري_I-SMISC _NM :_PX dحَدَّثَنَا_PSTV+PRO
الْحُمَيْدِيُّ_B-PER عَبْدُ_I-PER اللَّهِ_I-PER بْنُ_I-PER الزُّبَيْرِ_I-PER قَالَ_PSTV
حَدَّثَنَا_PSTV+PRO سُفْيَانُ_B-PER قَالَ_PSTV حَدَّثَنَا_PSTV+PRO
يَحْيَى_B-PER بْنُ_I-PER سَعِيدٍ_I-PER الْأَنْصَارِيُّ_I-PER قَالَ_PSTV
أَخْبَرَنِي_PSTV+PRO مُحَمَّدُ_B-PER بْنُ_I-PER إِبْرَاهِيمَ_I-PER التَّيْمِيُّ_I-PER
أَنَّهُ_ACC+PRO سَمِعَ_PSTV عَلْقَمَةَ_B-PER بْنَ_I-PER وَقَّاصٍ_I-PER
اللَّيْثِيَّ_I-PER يَقُولُ_PRSV سَمِعْتُ_PSTV عُمَرَ_B-PER بْنَ_I-PER
الْخَطَّابِ_I-PER رَضِيَ_PSTV اللَّهُ_ALLAH عَنْهُ_P+PRO عَلَى_P
الْمِنْبَرِ_D+SMN قَالَ_PSTV سَمِعْتُ_PSTV رَسُولَ_B-PROPH اللَّهِ_I-PROPH
صَلَّى_PSTV اللَّهُ_ALLAH عَلَيْهِ_P+PRO وَسَلَّمَ_C+PSTV يَقُولُ_PRSV
إِنَّمَا_ACC+PREV الْأَعْمَالُ_D+PIN بِالنِّيَّاتِ_P+D+PFN وَإِنَّمَا_C+ACC+PREV
كَانَتْ_PSTV فَمَنْ_C+REL نَوَى_PSTV مَا_REL امْرِئٍ_SMN لِكُلِّ_P+NQ
إِلَى_P أَوْ_C إِلَى_P دُنْيَا_SFN يُصِيبُهَا_PRSV+PRO هِجْرَتُهُ_SFN+PRO
إِلَى_P مَا_REL فَهِجْرَتُهُ_C+SFN+PRO يَنْكِحُهَا_PRSV+PRO امْرَأَةٍ_SFN
هَاجَرَ_PSTV إِلَيْهِ_P+PRO

Figure 7: The sample output of A-POS

<START:MISC> صحيح البخاري <END> : حَدَّثَنَا <START:PER> الْحُمَيْدِيُّ
قَالَ حَدَّثَنَا <END><START:PER> سُفْيَانُ <END> قَالَ حَدَّثَنَا <START:PER> عَبْدُ اللَّهِ بْنُ الزُّبَيْرِ
قَالَ أَخْبَرَنِي <END><START:PER> يَحْيَى بْنُ سَعِيدٍ الْأَنْصَارِيُّ
أَنَّهُ سَمِعَ <END><START:PER> مُحَمَّدُ بْنُ إِبْرَاهِيمَ التَّيْمِيُّ <END><START:PER>
عَلْقَمَةَ بْنَ وَقَّاصٍ اللَّيْثِيَّ <END><START:PER> يَقُولُ سَمِعْتُ <END><START:PER>
عُمَرَ بْنَ الْخَطَّابِ <END><START:ALLAH> رَضِيَ اللَّهُ <END> سَمِعْتُ
عَلَى الْمِنْبَرِ قَالَ <START:PROPH> رَسُولَ اللَّهِ <END><START:ALLAH> صَلَّى اللَّهُ <END>
عَلَيْهِ وَسَلَّمَ يَقُولُ إِنَّمَا الْأَعْمَالُ بِالنِّيَّاتِ وَإِنَّمَا لِكُلِّ امْرِئٍ مَا نَوَى فَمَنْ كَانَتْ هِجْرَتُهُ إِلَى
دُنْيَا يُصِيبُهَا أَوْ إِلَى امْرَأَةٍ يَنْكِحُهَا فَهِجْرَتُهُ إِلَى مَا هَاجَرَ إِلَيْهِ

Figure 8: The sample output of A-NER

The next step is *Sanad* Hadith Graph Construction. This step adopted one of the approaches for Graph modeling from [29] that described as follow:

1. Describe the Model in Terms of the Application's Needs.
2. Nodes for Things, Relationships for Structure
3. Fine-Grained versus Generic Relationships
4. Model Facts as Nodes
5. Represent Complex Value Types as Nodes
6. Iterative and Incremental Development

This Graph construction step employed Python programming and Neo4j Graph database. The new process was done to outfitted the Narrators node with their other attributes from the Narrator taken from the database Narrator (*Rijal Al-Hadits*) includes attributes: Full Name, *Kuniyah*, *Laqob*, Generation, Year of Birth, Year of Death, and Grade. Table 2 shows the sample of attributes for Graph Narrator node with the name عمر بن الخطاب. Furthermore, the output is the Graph model of *Sanad* (Narrator chain). Figure 9 shows the specimen of one chain of *Sanad* Graph.

Table 2: Sample of Narrator Graph Attributes.

Name	عمر بن الخطاب
Full Name	'Umar ibn al-Khattab
Kuniyah	أبو حفص
Laqob	الفاروق
Generation	1st
Year Of Birth	41 BH
Year Of Death	23 AH
Grade	Companion (صحابة)

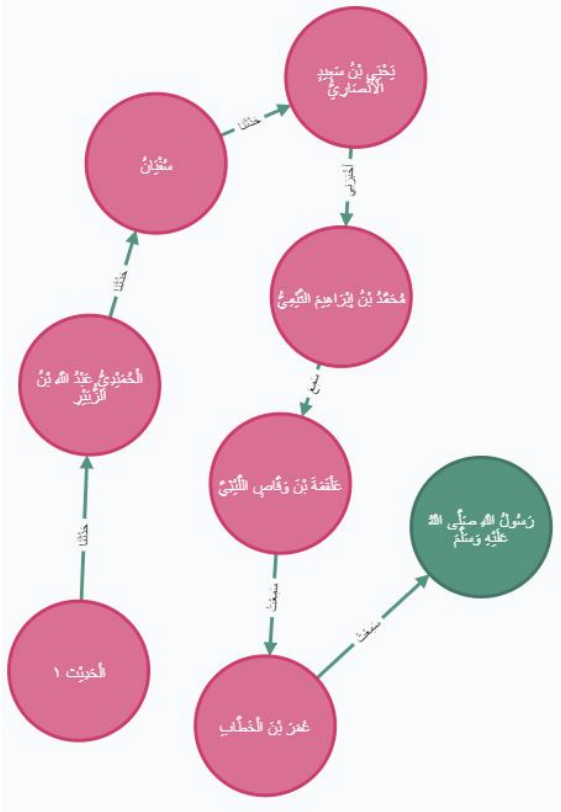


Figure 9: The sample output of A-NER

The challenge in defining the relationship between each Narrator on the Hadith was on the condition when Hadith had more than one *Sanad* lane. The previous sample or raw Hadith text on Figure 5 is a simple Hadith text with one *Sanad* lane. Figure 10 shows the Hadith with three *Sanad* lanes. It was needed a mechanism to define which the Narrator was connected, and which was not connected. This study was utilizing a machine learning model to determine the *Sanad* Narrator connection. The input of model was the value of a Generation of Narrator, Year of Birth, Year of Death, Number of appearance in the Hadith, Position in the Hadith that extracted manually from Hadith *Sanad*. The output of the model was the status of Narrator Connection. Table 3 shows the specimen of the data used to train the machine learning model. Overall 263 Narrator combination used extracted from 17 Hadith *Sanad*.


صحيح البخاري ٥: حَدَّثَنَا عَبْدَانُ قَالَ أَخْبَرَنَا عَبْدُ اللَّهِ قَالَ أَخْبَرَنَا يُونُسُ عَنْ الزُّهْرِيِّ ح و حَدَّثَنَا بِشْرُ بْنُ مُحَمَّدٍ قَالَ أَخْبَرَنَا عَبْدُ اللَّهِ قَالَ أَخْبَرَنَا يُونُسُ وَمَعْمَرٌ عَنْ الزُّهْرِيِّ نَحْوَهُ قَالَ أَخْبَرَنِي عُبَيْدُ اللَّهِ بْنُ عَبْدِ اللَّهِ عَنْ ابْنِ عَبَّاسٍ قَالَ كَانَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ أَجْوَدَ النَّاسِ وَكَانَ أَجْوَدُ مَا يَكُونُ فِي رَمَضَانَ حِينَ يَلْقَاهُ جِبْرِيلُ وَكَانَ يَلْقَاهُ فِي كُلِّ لَيْلَةٍ مِنْ رَمَضَانَ فَيُدَارِسُهُ الْقُرْآنَ فَلَرَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ أَجْوَدُ بِالْخَيْرِ مِنْ الرِّيحِ الْمُرْسَلَةِ

Figure 10: The Arabic Hadith Text with three lanes of *Sanad*

Table 3: Sample of Training Data for *Sanad* Connection Machine Learning Model

Gen 1	Birth 1	Death 1	Appr 1	Pos 1	Gen 2	Birth 2	Death 2	Appr 2	Pos 2	C
10	0	219	1	1	2	10	75	1	5	0
10	0	219	1	1	0	-41	23	1	6	0
8	107	196	1	2	5	-70	144	1	3	1
8	107	196	1	2	4	0	120	1	4	0
10	142	226	1	1	3	94	104	1	8	0
10	142	226	1	1	0	-16	78	1	9	0
7	94	175	1	2	6	0	144	1	3	1
7	94	175	1	2	4	51	124	2	4	0

The model built by utilized and then evaluated four nonlinear algorithms, that are:

- Classification and Regression Trees (CART)
- Support Vector Machines (SVM)
- Gaussian Naive Bayes (NB)
- K-Nearest Neighbors (KNN)

Another way to improve the performance of algorithms is by using ensemble methods. Four different ensemble machine learning algorithms, two boosting and two bagging methods were utilized and evaluated that are:

- Boosting Methods: AdaBoost (AB) and Gradient Boosting (GBM)

- Bagging Methods: Random Forests (RF) and Extra Trees (ET).

Figure 11 visualizes mean and standard deviation for each non-linear algorithm accuracy prediction.

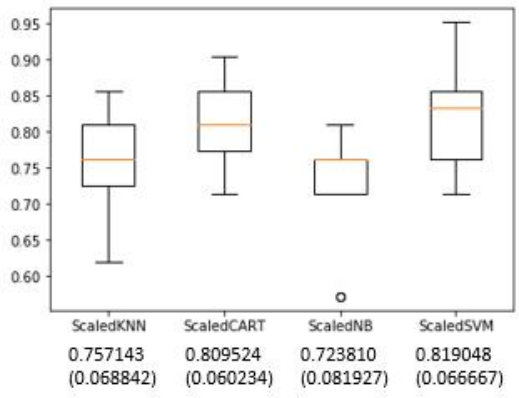


Figure 11: *Sanad* Link Prediction – Non Linear Algorithm Comparison

Figure 12 visualizes mean and standard deviation for each ensemble algorithm accuracy prediction.

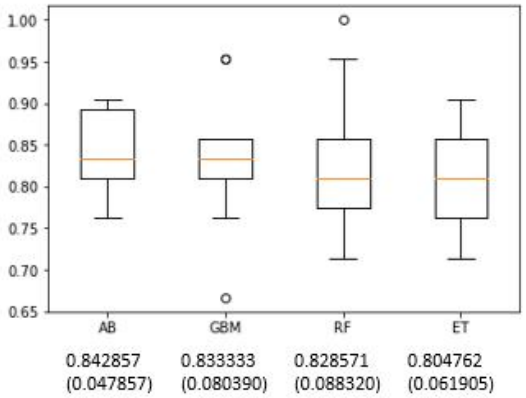


Figure 12.:*Sanad* Link Prediction – Ensemble Algorithm Comparison

The prediction shows that SVM and GBM have the best accuracy mean with values are 0.821795 and 0.821795. The final evaluation values for each algorithm employed the confusion matrix shown in Table 4.

Table 4.:Confusion Matrix Result

Algo-rithm		Precision	Recall	F1-Score	Support
SVM	0	0.81	0.97	0.89	36
	1	0.90	0.53	0.67	17
	micro avg	0.83	0.83	0.83	53
	macro avg	0.86	0.75	0.78	53
	weighted avg	0.84	0.83	0.82	53
GBM	0	0.88	1.00	0.94	36
	1	1.00	0.71	0.83	17
	micro avg	0.91	0.91	0.91	53
	macro avg	0.94	0.94	0.88	53
	weighted avg	0.92	0.92	0.90	53

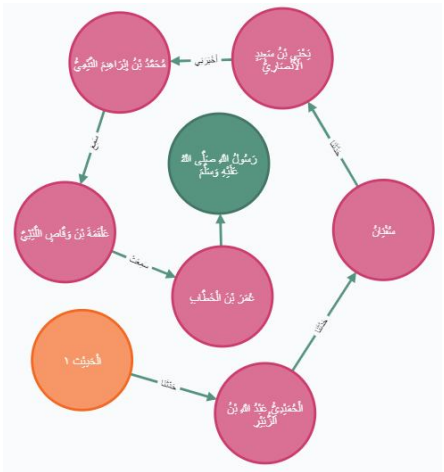
4.2 Result

Utilizing Python and Neo4j with a chipper query language, the final *Sanad* Graph results shown in Figure 9. The verification process for the *Sanad* Graph developed was done by verifying all chains of the *Sanad* Graph with the original Hadith texts by generating Evaluation Question as shown in Table 5.

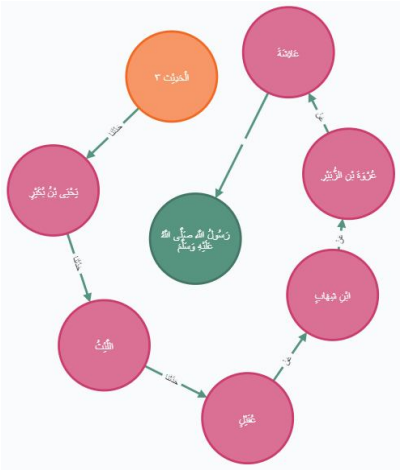
Table 5: Evaluation Question

Id	Research Question
Q1	Who are the Narrators in *Sanad* Hadith 1?
	MATCH p=(m { name: 'الْحَدِيْث ١' })-[r*..10]->(n) WHERE ALL(rel in r WHERE rel.Hadith='الْحَدِيْث ١') RETURN p
Q2	Who are the Narrators in *Sanad* Hadith 2?
	MATCH p=(m { name: 'الْحَدِيْث ٢' })-[r*..10]->(n) WHERE ALL(rel in r WHERE rel.Hadith='الْحَدِيْث ٢') RETURN p
Q3	Who are the Narrators in *Sanad* Hadith 3?
	MATCH p=(m { name: 'الْحَدِيْث ٣' })-[r*..10]->(n) WHERE ALL(rel in r WHERE rel.Hadith='الْحَدِيْث ٣') RETURN p
Q4	Which Hadith did 'أَنَس' participate in its *Sanad*?
	MATCH p=(m { name: 'أَنَس' })<-[r*..10]-(n) RETURN p
Q5	Which Hadith did 'الزُّهْرِيّ' participate in its Sanad?
	MATCH p=(m { name: 'الزُّهْرِيّ' })<-[r*..10]-(n) RETURN p
Q6	Which Hadith did 'عَائِشَة' participate in its *Sanad*?
	MATCH p=(m { name: 'عَائِشَة' })<-[r*..10]-(n) RETURN p

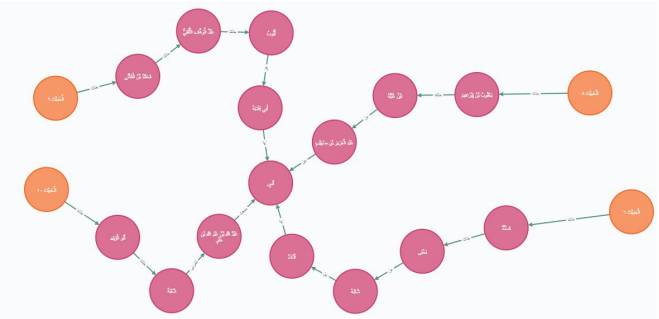
Q1, Q2, Q3 evaluated the *Sanad* chains for Hadith 1, 2, 3 and also done for Hadith number 4 until 8 to verify the correctness of the Graph resulted. Q4, Q5, Q6 repeated for all Narrator names in the Graph.



(i)



(ii)



(iii)

Figure 13: Result of Q1, Q2, and Q4

Figure 13 sequentially show the outcomes of Q1, Q2, and Q4. Q4 result show visually that in أَنَس involved as a Narrator for Hadith 6, 8, 9 and 10.

5. CONCLUSION

This study set out to proposed a novel graph-based representation for Hadith *Sanad*. First, the candidate of graph node and relation were extracted automatically from the example of thirty raw Arabic Hadith text collected manually from http://qaalarasulallah.com/. Second, the Arabic Part of Speech (A-POS) and Arabic Named Entity Recognition (A-NER) that adopted out of the effort of [33] exploited to do that extraction. Third, SVM and GBM algorithm employed to develop the new machine learning model for the Sanad Hadith Graph construction. Fourth, the *Sanad* Hadith Graph built utilizing Python and Neo4j with a chipper query language. The final step was the *Sanad* Hadith Graph verification. The result shows that the *Sanad* Graph of Hadith Narrator built successfully and already verified the correctness compare with the original Hadith text.

REFERENCES

[1] R. Baru, S. Hadzrullathfi, S. Omar, and B. Ibrahim, "Identifying False Hadith Guidelines," *Malaysian J. Islam. Stud.*, vol. 1, pp. 62–73, 2017.

[2] S. R. Mohammad Najib, N. Abd Rahman, N. Kamal Ismail, N. Alias, Z. Mohamed Nor, and M. N. Alias, "Comparative Study of Machine Learning Approach on Malay Translated Hadith Text Classification based on Sanad," *MATEC Web Conf.*, vol. 135, p. 00066, 2017. https://doi.org/10.1051/matecconf/201713500066

[3] N. A. P. Rostam and N. H. A. H. Malim, "Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting," *J. King Saud Univ. - Comput. Inf. Sci.*, Mar. 2019.

[4] Muahammad Khurram Khan and M.Alginahi Yaser, "The Holy Quran Digitization: Challenges and Concerns," *Life Sci. J.*, vol. 10, no. 2, pp. 156–164, 2013.

[5] Ayub, "Matn Criticism and Its Role in The Evaluation of Hadith Authenticity," *(International J. Islam. Stud. Humanit.*, no. March, pp. 1–4, 2018. https://doi.org/10.26555/ijish.v1i1.136

[6] A. Mahmood, H. Ullah, F. K., M. Ramzan, and M. Ilyas, "A Multilingual Datasets Repository of the Hadith Content," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, 2018.

[7] E. T. Luthfi, N. Suryana, and A. S. H. Basari, "Digital Hadith Authentication: a Literature Review and Analysis," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 15, 2018.

[8] M. Najeeb, A. Abdelkader, M. Al-Zghoul, and A. Osman, "A Lexicon for Hadith Science Based on a Corpus," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 2, pp. 1336–1340, 2015.

[9] 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. https://doi.org/10.1007/s10462-016-9458-x

[10] A. Zakaria, *Pokok-pokok Ilmu Musthalah Hadits*, Pertama. Garut: Ibn Azka Press, 2014.

[11] S. I. Hyder and S. G. Ahmed, "Towards a Database Oriented Hadith Research Using Relational, Algorithmic and Data-Warehousing Techniques," *Res. Q. J. Shaikh Zayed Islam. Cent. Islam. Arab. Stud.*, vol. 19, no. March, pp. 14–32, 2008.

[12] N. N. Amirah, T. M. Rahim, Z. Mabni, H. M. Hanum, and N. A. Rahman, "A Malay Hadith translated document retrieval using parallel Latent Semantic Indexing (LSI)," *2016 3rd Int. Conf. Inf. Retr. Knowl. Manag. CAMP 2016 - Conf. Proc.*, pp. 118–123, 2017.

[13] I. A. Khan, *Authentication of Hadith: Redefining Criteria*. International Institute of Islamic Thought, 2012.

[14] H. U. Next, "The Science of Hadith," 2007. [Online]. Available: http://www.islamic-awareness.org/Hadith/Ulum/hadsciences.html. [Accessed: 07-Oct-2017].

[15] Damanhuri, "The Hadith of Rajab Fasting in a Simultaneous Research," *Int. J. Islam. Stud.*, vol. 3, no. 1, pp. 221–248, 2016. https://doi.org/10.20859/jar.v3i1.83

[16] M. Zein, *Ilmu Memahami Hadits Nabi Cara Praktis Menguasai Ulumul Hadits & Mustholah Hadits*, I. Yogyakarta: Pustaka Pesantren, 2013.

[17] N. Alias, N. A. Rahman, N. K. Ismail, Z. M. Nor, and M. N. Alias, "Graph-based text representation for Malay translated hadith text," in *2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2016, pp. 60–66.

[18] H. Maraoui, K. Haddar, and L. Romary, "Encoding Prototype of Al-Hadith Al-Shareef in TEI," in *Communications in Computer and Information Science*, vol. 782, Springer, Cham, 2018, pp. 217–229.

[19] H. A. Al-Sanasleh and B. H. Hammo, "Building Domain Ontology: Experiences in Developing the Prophetic Ontology Form Quran and Hadith," in *2017 International Conference on New Trends in Computing Sciences (ICTCS)*, 2017, vol. 2018-Janua, pp. 223–228. https://doi.org/10.1109/ICTCS.2017.35

[20] B. Fairouz and T. Nora, "Computational Ontologies for a Semantic Representation of the Islamic Knowledge," Springer, Cham, 2018, pp. 37–46.

[21] A. H. Jaafar and N. C. Pa, "Hadith Commentary Repository: An Ontological Approach," in *Proceedings of the 6th International Conference on Computing and Informatics*, 2017, no. 167, pp. 191–198.

[22] A. Al-Rumkhani, M. Al-Razgan, and A. Al-Faris, "TibbOnto: Knowledge Representation of Prophet Medicine (Tibb Al-Nabawi)," *Procedia Comput. Sci.*, vol. 82, no. March, pp. 138–142, 2016. https://doi.org/10.1016/j.procs.2016.04.021

[23] S. Mohamed Osman Hassan and E. Atwell, "Design and Implementing Of Multilingual Hadith Corpus," *Int. J. Recent Res. Soc. Sci. Humanit.*, vol. 3, no. 2, pp. 100–104, 2016.

[24] Rebhi S. Baraka and Yehya M. Dalloul, "Building Hadith Ontology to Support the Authenticity of Isnad," *Int. J. Islam. Appl. Comput. Sci. Technol.*, vol. 2, no. 1, pp. 25–39, 2014.

[25] A. J. El-Massry, "Ontology-based Approach to Support Semantic Search in Hadith (Zakat Domain)," Gaza, 2017.

[26] M. Malhotra and T. R. G. Nair, "Evolution of Knowledge Representation and Retrieval Techniques," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 7, pp. 18–28, Jun. 2015.

[27] R. kumar Kaliyar, "Graph databases: A survey," in *International Conference on Computing, Communication & Automation*, 2015, pp. 785–790.

[28] J. Pokorný, "Graph Databases: Their Power and Limitations," in *Lecture Notes in Computer Science*, vol. 8104, 2015, pp. 58–69.

[29] I. Robinson, J. Webber, and E. Eifrem, *Graph Databases*. O'Reilly Media, 2013.

[30] R. McNeely, "Data Modeling with Graph Databases," 2014.

[31] J. Pujara and S. Singh, "Mining Knowledge Graphs From Text," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18*, 2018, no. 3, pp. 789–790. https://doi.org/10.1145/3159652.3162011

[32] F. Shahzad, "Development of Innovative Islamic Web Applications," *Int. J. Ind. Manuf. Eng.*, vol. 10, no. 9, pp. 3134–3139, 2016.

[33] A. A. Freihat, G. Bella, H. Mubarak, and F. Giunchiglia, "A single-model approach for Arabic segmentation, POS tagging, and named entity recognition," *2nd Int. Conf. Nat. Lang. Speech Process. ICNLSP 2018*, pp. 1–8, 2018.

[34] T. Hariguna, MT. Lai, CW. Hung, and SC. Chen, "Understanding information system quality on public e-government service intention: an empirical study", Int. J. Innovation and Sustainable Development, Vol. 11, No. 2/3, 271-290, 2017. https://doi.org/10.1504/IJISD.2017.083290

[35] U. Rahardja, T. Hariguna, & Q. Aini, "Understanding the Impact of Determinants in Game Learning Acceptance: An Empirical Study", International Journal of Education and Practice, 7(3), 136–145, 2019. http://doi.org/10.18488/journal.61.2019.73.136.145

[36] B. Buvaneswari and T. K. Reddy, "ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial Recognition System," Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 1, pp. 12–17, 2019. https://doi.org/10.30534/ijatcse/2019/03812019

[37] S. B. B, R. S. Rajkumar, and B. F. Ibrahim, "Service Profile based Ontological System for Selection and Ranking of Business Process Web Services," Int. J. Adv. Trends Comput. Sci. Eng., vol. 8, no. 1, pp. 18–22, 2019. https://doi.org/10.30534/ijatcse/2019/04812019