SEMANTIC GRAPH KNOWLEDGE REPRESENTATION FOR AL-QURAN VERSES BASED ON WORD DEPENDENCIES

Muhammad Muhtadi Mohamad Khazani[1]*, Hassan Mohamed[2], Tengku Mohd Tengku Sembok[3]*, Nurhafizah Moziyana Mohd Yusop[4], Sharyar Wani[5], Yonis Gulzar[6], Mohd Hazali Mohamed Halip[7], Syahaneim Marzukhi[8], Zahri Yunos[9]

[1,2,3,7,8]Cyber Security and Digital Industrial Revolution Centre, National Defence University of Malaysia, Kuala Lumpur, Malaysia

[4]Faculty of Defence Science and Technology, National Defence University of Malaysia, Kuala Lumpur, Malaysia

[5]Department of Computer Science, Kulliyyah of Information & Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia

[6]Department of Management Information Systems, College of Business Administration, King Faisal University, Al-Ahsa, Saudi Arabia

[9]CyberSecurity Malaysia, Selangor, Malaysia

Email: mmad.muhtadi@gmail.com[1]*(corresponding author), hassan@upnm.edu.my[2], tmts@upnm.edu.my[3]*(corresponding author), moziyana@upnm.edu.my[4], sharyarwani@iium.edu.my[5], ygulzar@kfu.edu.sa[6], hazali@upnm.edu.my[7], syahaneim@upnm.edu.my[8], zahri@cybersecurity.my[9]

ABSTRACT

Semantic approaches present an efficient, detailed and easily understandable representation of knowledge from documents. Al-Quran contains a vast amount of knowledge that needs appropriate knowledge extraction. A semantic based approach can help in designing an efficient and explainable knowledge representation model for Al-Quran. This research aims to propose a semantic-graph knowledge representation model for verses of Al-Quran based on word dependencies. These features are used in the proposed knowledge representation model allowing the semantic graph matching to improve Al-Quran search applications' accuracy. The proposed knowledge representation model is essentially a formalism for generating a semantic graph representation of Quranic verses, which can be applied for knowledge base construction for other applications such as information retrieval system. A set of rules called Semantic Dependency Triple Rules are defined to be mapped into the semantic graph representing the verse's logic. The rules translate word dependencies and other NLP metadata into a triple form that holds logical information. The proposed model has been tested with English translation of Al-Quran on a document retrieval prototype The basic system has been enhanced with anaphoric pronouns correction, which has shown improvement in retrieval performance. The results have been compared with a closely related system and evaluated on the accuracy of the document retrieval in Precision, Recall and F-score measurements. The proposed model has achieved 65%, 60% and 62.4% for the measurements, respectively. It has also improved the overall accuracy of previous system by 43.8%.

Keywords: Al-Quran, Knowledge Representation, Document Retrieval, Association Rules, Word Dependencies

1.0 INTRODUCTION

Al-Quran is the most essential resource for Islamic knowledge as well as the Arabic language [1]. It has been continuously referred by Muslims and non-Muslims for its immense knowledge, which includes jurisprudence, moral values, historical anecdotes, and scientific matters. The Islamic foundations are based on two primary sources – Al-Quran and Hadith. Hence, understanding Al-Quran becomes essential in Islam. In the past, extracting the knowledge from Al-Quran and understanding its content was entrusted to many Islamic scholars. Nowadays, Al-Quran has been digitalised like other documents that were traditionally written on paper. The emergence of artificial intelligence has significantly made the knowledge extraction easier and efficient. Knowledge representation and reasoning is a well-known topic in artificial intelligence that concerns the symbolic representation of knowledge and the reasoning process for knowledge manipulation. Furthermore, artificial intelligence is one of the driving forces in the fourth industrial revolution.

132

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

The most popular knowledge representation methods in general are logic, semantic networks, frame, and production rule. The logic-based approach has a precisely defined syntax and semantics. Thus, it serves as a reliable reasoning algorithm. Syntax refers to a language's structure, whereas semantics refer to the meaning and interpretation [2]. Logic has firm rules that deal with different conditions [2], [3]. Logical representation refers to either propositional logic or predicate logic. On the other hand, a semantic network represents knowledge in graph form, which comprises nodes and arcs [2], [4]. A node can represent an object. An arc between nodes represents a relationship between objects. Semantic network representation is computationally expensive, but it is good at representing meanings [2]. Frame language is another knowledge representation method that deals with classes and objects [2], [3]. The frame is made up of slot and facet. A slot represents an attribute of the object specified by the frame. Each slot has a variety of detailed information called facets. Frame representation can be visualised easily, but it is difficult for the reasoning process. Production rule represents knowledge using a causal structure such as "If A, then B" [2], [4]. A rule-based system can acquire results through reasoning by manipulating initial information. Computation can be less complex since rules can be altered independently [2]. The research questions behind these knowledge representations techniques are how to adapt them for representing the knowledge of Al-Quran, since the verse of Al-Quran is complex. For example, a verse may comprise of several sentences that having many anaphoric references.

Various knowledge representation methods have been applied to represent the knowledge of Al-Quran. These knowledge bases have been used by other applications such as information retrieval and question answering systems [5]–[9]. Common ontology-based knowledge representation models for Al-Quran focuses on concepts, rather than semantic relations that reside within the Quranic text [10]. Furthermore, ontological approaches to represent knowledge in Al-Quran has been used in [6], [8], [11]–[13] by extracting entities or concepts in Al-Quran. Even this approach looks good in representing concepts in the verses of Al-Quran, it seems no attempts to represent knowledge in Al-Quran beyond the context since ontology emphasis on concept hierarchy. On the other hand, semantic graph representation has shown great potential because it allows the logical semantic approach to add semantic elements of the Quranic verses. Previous works have made effort to establish semantic graph knowledge representation through relations between semantic units in Al-Quran [10], [14], [15]. However, the work relied on phrase structure of parsed Quranic text which only considering the syntactic structure of the sentences. The semantic approach can provide better knowledge representation for Quranic text because it is taken from natural language syntactic and semantic such as word dependencies in a sentence. Therefore, this paper aims to propose a semantic-based knowledge representation model for verses of Al-Quran that uses word dependencies which cater on both syntactic and semantic of a sentence of the parsed Quranic text. A good knowledge representation model for Al-Quran will surely help people worldwide to extract information from Al-Quran and understand it better [16].

2.0 LITERATURE REVIEW

In this section, related works in knowledge representation system and knowledge retrieval system for Al-Quran are reviewed. Methods and approaches that have been used in previous works are highlighted to get the general idea of current research and development in the domain of Al-Quran. Challenges and area for improvement on the previous works are identified for further research.

2.1 Knowledge Representation Systems for Al-Quran

Al-Quran is a unique and complex corpus that contains vast knowledge to be explored. Thus, the knowledge that resides within Al-Quran needs to be efficiently represented for further use to get the most accurate retrieval or processing. Most of the knowledge representations are based on ontology [5], [6], [8]. Ontology is a formal, explicit specification of a shared conceptualisation [17]. It represents a collection of concepts of a particular domain such as medical, manufacturing, education as well as Al-Quran [6], [18], [19]. Other than ontology, the logic and linguistic approach uses first-order predicate logic to develop the consistent logico-linguistic formalism for Al-Quran [10]. Noun-verb structures linked to an event description, events, and role analysis become the primary focus of syntactic analysis of the text, and they are translated to predicate-subject forms. This formalism has been applied to construct a knowledge base and tested on a list of questions pertaining to knowledge in Al-Quran.

Ontology has been applied to the knowledge about stories of prophets in Al-Quran [6] to provide more expressive representations of Quran relations in terms of rules. The concepts and semantic relations in Al-Quran are extracted using linguistic pattern-based schemes and associations rules. Apriori algorithm was applied to generate a set of rules for concept extraction. A study has reviewed the existing applications on Al Quran ontology and found that the most common datasets related to Al-Quran are the Quranic topics (QT), Arabic Quran Corpus (AQC), and QuranA [20]. However, most of the Quran ontologies do not include semantic tags for all Quranic verses. Ontology-based

133

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

approaches also face common problem of high cost in effort, time and money for developing and maintaining an ontology[6], [21]. An effort has been made to develop a semi-automatic system that reuses existing ontologies and adapting them to suit current needs. It is a collaborative efforts between knowledge engineers and domain experts [21].

As there is a massive amount of data that needs to be processed, it is important to automate the knowledge extraction procedures. Therefore, a knowledge representation model based on rules is a good approach to extract knowledge from the data and store it in a knowledge base. The rule-based knowledge base construction should cover various scenarios in term of sentence structures as well as its meanings [14].

2.2 Knowledge Retrieval Systems for Al-Quran

The goal of knowledge retrieval from a large volume of information is to provide user support for interpreting and obtaining adequate information [16]. The initial stage of the knowledge representation research had been carried out emphasising keyword-based methods [22]. These methods start from finding possible combinations of keywords that are relevant to the desired query. Therefore, most of the knowledge retrieval systems developed for Al-Quran are using keyword matching techniques [23]–[26]. Later, query expansion techniques are used to improve the accuracy rate [27]–[30]. Moreover, researchers used the semantic-based method where it helps to retrieve information using similarity concepts to match the contextual meaning of keyword [9], [25], [31], [32].

Traditional search applications that rely on keyword search face problems such as natural language ambiguity and reference reconciliation. As a result, keyword-based retrieval systems produce many irrelevant results [33]. Query expansion has been applied to include synonyms of words used in queries, improving the retrieval. This is achieved by using lexical databases such as WordNet.

An Arabic question answering system, QARAB, combined traditional information retrieval techniques with a sophisticated Natural Language Processing approach [26]. The system identified relevant text passages that potentially answer the questions. Additionally, a question answering system for Al-Quran called Al-Bayan, understand the semantics of Al-Quran using Support Vector Machine (SVM) by classifying questions such that the answers are ranked [34]. Question answering systems have used many techniques such as keyword matching, keyword expansion, SPARQL query, SVM Question classification & cosine similarity, Artificial Neural Network Verb classification and N-gram techniques to achieve better retrieval results [35].

In term of evaluation for knowledge retrieval systems, Precision and Recall are widely used as they measure the relevancy of the retrieved information [6], [33], [36]. F-score has also been used as it is the harmonic mean of Precision and Recall, which would show the retrieval accuracy of the systems [6]. Although Precision and Recall are the standards, other evaluation metrics have been used as well such as First Hit Success, First Answer Reciprocal Rank and Total Reciprocal Rank [37], [38].

3.0 METHODOLOGY

In this section, conceptual framework for the proposed model and its implementation are presented. Each component of the conceptual framework is briefly described. The formalism for generating semantic graph representation which is also described. The experimental setup consisting of test dataset, query processing and evaluation metrics are also explained.

3.1 Conceptual Framework

A conceptual framework is important for the proposed model to identify every integral process in order to produce desirable output from an input. Fig. 1 illustrates the conceptual framework for the proposed knowledge representation model and question answering prototype. This framework can be divided into two components: knowledge base construction and question answering system.

134

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021
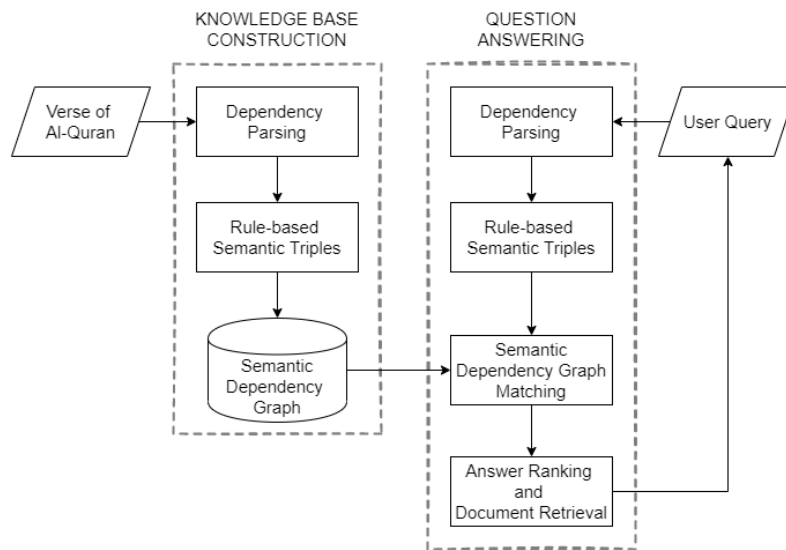
Fig. 1: Conceptual framework of the proposed knowledge representation model and question answering

3.1.1   Knowledge Base Construction

The knowledge base construction has two main processes before the generation of a knowledge graph database. In our knowledge base construction, the system receives verses of English translated Al-Quran as an input document. Then the verses are parsed using dependency parsing to get word dependencies, as well as other metadata. Using this information, the system will create semantic graphs after matching with a set of rules. These semantic graphs are then populated into a graph database to become a knowledge base. The semantic graphs in this knowledge base represent the meaning of the Quranic verses.

3.1.1.1   Dependency Parsing

The Quranic verses are parsed to obtain the grammatical units of a sentence and the dependency relations between words. Dependency parsing provides the relation in the sense that it is comparable to the meanings of the sentence. Dependency parsing yields word dependencies, consisting of governor words, dependent words and grammatical relations between words [39], [40]. The dependencies generally comprise subjects, objects and modifiers. Every word in a sentence is classified into one of these types depending on the sentence structure. On top of that, other NLP metadata such as lemmatisation, Part-of-Speech (POS) tags, Named-entity Recognition, and co-reference [41] can also be used in building the semantic graph. This information plays an essential role in improving knowledge representation and retrieval by putting words into the appropriate context as an instance. Other metadata such as verse number, chapter name and number of occurrences are also collected. This information is used to match relevant documents from which the knowledge is located. Additionally, the dependency parsing yields a dependency parse tree for every sentence. The parse tree is crucial for the analysis of syntax and semantics for selecting the appropriate rules (Semantic Dependency Triple Rules) before the final semantic graph of a sentence is generated.

3.1.1.2   Rule-based Semantic Triples

Semantic triples are a set of three types of information that signify semantics in subject, predicate, and object [42]. The generated semantic triples in this framework include all kinds of semantic relations representing meanings within a sentence. For example, the triples store predicates as the semantic relations between subjects and objects. The triples also store semantic relations between concepts that look like relations in ontology. Therefore, ruled-based semantic triples in this framework focus on representing the meanings carried by a sentence. These semantic triples can be represented in a graphical form. A triple can be visualised as a graph of two nodes and an arc connecting them [43]. The nodes represent a subject and an object, while the arc (directed from the subject to the object) represents the predicate. The graphical representation is called the Semantic Dependency Graph, is stored in a graph database as a knowledge base. The knowledge base will have the proposed semantic-based knowledge representation of Al-Quran.

135

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

3.1.2   Question Answering

The purpose of the question answering system in this framework is to evaluate the applicability of the proposed semantic-based knowledge representation model. Therefore, the effectiveness of the representational semantic of Al-Quran can be measured based on retrieval performance. The question answering component goes through the same process as knowledge base construction, except it, does not store semantic graph into the knowledge base. Instead, the semantic graphs built from queries are matched into the graphs in the knowledge base. The matching process returns a set of answers with their own rank and the source of relevant answers from documents. Hence, it is able to provide both answer and relevant document.

3.1.2.1   Semantic Dependency Graph matching

The semantic triples of the query are matched with the Semantic Dependency Graph in the knowledge base. The matching algorithm examines any similar nodes and relationships in the knowledge base that produce a list of similarly matched graphs. The system determines possible answers from the matched graphs and returns a list of candidate answers to the user.

3.1.2.2   Answer Ranking and Document Retrieval

In the matching process, the matching nodes and the relationship between the Semantic Dependency Graph of the query and each Semantic Dependency Graph of the answer candidates are counted. Hence, the ranks of the answer candidates are determined by the most match counts after performing the count sorting. The highest rank is assumed to be the most accurate answer. In addition, the relevant documents of the returned answers are provided. The relevant documents can be retrieved as the verse and chapter of the matched answer and graph. In this paper, the experiment focuses more on document retrieval.

3.2   Formulation of Semantic Dependency Triple Rules

A triple is usually a set of three data entities stored as subject-predicate-object [42]. However, in the analysis of Quranic verses, the semantic of the verses are listed to be arranged in triples form to store meanings in terms of subject-predicate-object with other combinations of data such as lemma, anaphoric reference, part of speech and named entity recognition. Since a Quranic verse may comprise several sentences, each Quranic verse is split into separate sentences to extract the meanings and transform them into a triple. Table 1 shows the extracted meanings and their triples for Surah Ali Imran, Verse 35. A triple uses a keyword instead of multiple words that form a phrase. These keywords are obtained from the lemma of the verb in the sentence. This analysis uses Stanford CoreNLP parse trees, dependencies, and other NLP annotations. The parse tree shows syntax structure with grammatical relations or dependencies between headwords and dependent words. Fig. 2 and Fig. 3 present a part of dependency parse tree of Verse 35 of Surah Ali Imran. Additionally, Table 2 shows NLP metadata that is gathered to be used later in rule development.

Table 1: List of meanings and triples for sentences of Verse 35 in Surah Ali Imran

Surah Ali Imran, Verse 35		
Remember when Imran's wife prayed to her Lord saying, "I have made a vow to dedicate to Your service whatever is in my womb. *Lord, accept it from me. You are All-hearing and All-knowing*".		
Sentences	**Meanings**	**Triples**
Lord, accept it from me.	I asked Lord	(I, ask, Lord)
	I asked to accept	(I, ask_to, accept)
	Lord accept it	(Lord, accept, it)
	Lord accept from me	(Lord, accept_from, me)
You are All-hearing and All-knowing".	You are All-hearing	(You, is, All-hearing)
	You are All-knowing	(You, is, All-knowing)

136

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021
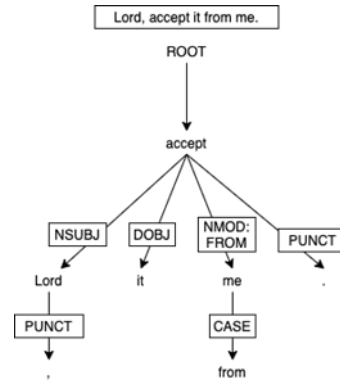
Fig. 2: Dependency parse tree obtained for the second sentence in Verse 35 of Surah Ali Imran
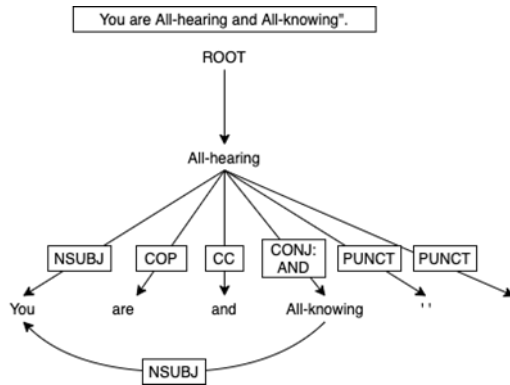

Fig. 3: Dependency parse tree obtained for the third and last sentence in Verse 35 of Surah Ali Imran

Table 2: NLP annotations for Verse 35 of Surah Ali Imran

Sentences	Tokens	Lemmas	POS Tags	Dependency
Lord, accept it from me.	Lord	Lord	NNP	root(ROOT-0, accept-3)
	,	,	,	nsubj(accept-3, Lord-1)
	accept	accept	VB	nmod:from(accept-3, me-6)
	it	it	PRP	punct(accept-3, .-7)
	from	from	IN	dobj(accept-3, it-4)
	me	I	PRP	case(me-6, from-5)
	.	.	.	punct(Lord-1, ,-2)
You are All-hearing and All-knowing".	You	you	PRP	root(ROOT-0, All-hearing-3)
	are	be	VBP	nsubj(All-hearing-3, You-1)
	All-hearing	all-hearing	JJ	cop(All-hearing-3, are-2)
	and	and	CC	punct(All-hearing-3, "-6)
	All-knowing	all-knowing	JJ	punct(All-hearing-3, .-7)
	"	"	"	cc(All-hearing-3, and-4)
	.	.	.	conj:and(All-hearing-3, All-knowing-5)
				nsubj(All-knowing-5, You-1)

Post syntactic and semantic analysis, the word dependencies and semantic triples can be matched. Based on the observation in the analysis, we found that different sentences possibly return similar word dependencies, which apply to similar semantic triples. Table 3 shows semantic triples that represent meanings in different sentences. The table shows that the semantic triples have a similar structure of subject-verb-object where each sentence has similar dependencies (dobj and nsubj) between words. We refer to this pattern a Direct Object pattern. This pattern repeatedly occurred in Al-Quran for different semantic triples with similar dependencies of dobj and nsubj. Further, by studying multiple parse trees from the verses of Al-Quran, we have identified several other patterns. These patterns become the basis of our proposed association rules.

137

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

Table 3: Word dependencies and POS tags of pattern subject-verb-object

Semantic Triples	Relevant Word Dependencies	POS tags
(I, make, vow)	dobj(made, vow) nsubj(made, I)	I – PRP made – VBN vow – NN
(God, know, this)	dobj(knew, this) nsubj(knew, God)	God – NN knew – VBD this – DT
(Lord, accept, offer)	dobj(accepted, offer) nsubj(accepted, Lord)	Lord – NNP accepted – VBD offer – NN
(Zachariah, take, custody)	dobj(took, custody) nsubj(took, Zachariah)	Zachariah – NNP took – VBD custody – NN
(he, ask, her)	dobj(ask, her) nsubj(ask, He)	He – PRP ask – VB her – PRP

Based on the patterns identified in the analysis, the rules are developed. Other NLP metadata is also considered, such as POS tags and named entities in the rule development. POS tags and named entities are extensively used in defining the rules, whilst lemmas are mainly used as the keyword in a node (the name of the node). For example, for the Direct Object pattern, the word dependencies that are considered as important are direct object (dobj) and nominal subject (nsubj). This pattern has a head (word) as a common anchor for both dependencies. The word dependencies that are considered important are direct object (dobj) and nominal subject (nsubj). In this case, the head must be a verb; hence any verb tagged as VB, VBD, VBG, VBN, VBP or VBZ as its part of speech is applicable to this rule. Both dependent words for dobj and nsubj dependencies must be a noun or pronoun and tagged with POS tags NN, NNP, NNPS or PRP. For the semantic triples construction, the first data in the triples is the dependent word for nsubj dependency followed by the head word for both dobj and nsubj dependencies. The last data in the triples is the dependent word for dobj dependency. This information is formalised as antecedents in the rule of Direct Object Rule, as presented by Rule 1 to form a triple.

Rule 1: Direct Object Rule

IF dependency(word1, word2) = DIRECT OBJECT
AND dependency(word1, word3) = NOMINAL SUBJECT
AND part-of-speech(word1) = VERB
AND part-of-speech(word2) = NOUN
OR part-of-speech(word2) = PRONOUN
AND part-of-speech(word3) = NOUN
OR part-of-speech(word3) = PRONOUN
THEN triples(word3, WORD1, word2)

All rules are tested by applying them to various dedicated sentences and redefined to avoid triples misrepresentation.

Table *4* shows the most important rules for semantic triples generation according to linguistic patterns. These rules can be used in generating semantic triples and have been used for the whole Al-Quran in the context of the current research.

Table 4: Semantic Dependency Triples rules

Patterns	Rules
Copular	**IF** dependency(word1, word2) = COPULAR **AND** dependency(word1, word3) = NOMINAL SUBJECT **AND** part-of-speech(word1) = NOUN **OR** part-of-speech(word1) = ADJECTIVE **THEN** triples(word3, IS_A, word1) **OR** triples(word3, IS, word1)
Negation with Copular	**IF** dependency(word1, word2) = NEGATION **AND** dependency(word1, word3) = NOMINAL SUBJECT

	AND dependency(word1, word4) = COPULAR **AND** part-of-speech(word1) = NOUN **OR** part-of-speech(word1) = ADJECTIVE **THEN** triples(word3, IS_NOT_A, word1) **OR** triples(word3, IS_NOT, word1)
Adjectival Modifier	**IF** dependency(word1, word2) = ADJECTIVAL MODIFIER **AND** part-of-speech(word2) = ADJECTIVE **THEN** triples(word4, IS, word2)
Direct Object	**IF** dependency(word1, word2) = DIRECT OBJECT **AND** dependency(word1, word3) = NOMINAL SUBJECT **AND** part-of-speech(word1) = VERB **THEN** triples(word3, WORD1, word2)
Nominal Modifier with Verb Preposition	**IF** dependency(word1, word2) = NOMINAL MODIFIER:PREPOSITION **AND** dependency(word1, word3) = NOMINAL SUBJECT **AND** part-of-speech(word1) = VERB **THEN** triples(word3, WORD1_PREPOSITION, word2)
Nominal Modifier with Of Preposition	**IF** dependency(word1, word2) = NOMINAL MODIFIER:OF **AND** dependency(word1, word3) = NOMINAL SUBJECT **AND** part-of-speech(word1) = NOUN **THEN** triples(word3, OF, word2)
Nominal Modifier with Possessive	**IF** dependency(word1, word2) = NOMINAL MODIFIER:POSSESSIVE **THEN** triples(word1, OF, word2)
Conjunction for Nouns	**IF** dependency(word1, word2) = CONJUNCTION:AND **OR** dependency(word1, word2) = CONJUNCTION:OR **THEN** triples(word1, AND, word2) **OR** triples(word2, OR, word2)
Conjunction for Adjectives	**IF** dependency(word1, word2) = CONJUNCTIONS:AND **OR** dependency(word1, word2) = CONJUNCTIONS:OR **AND** dependency(word1, word3) = NOMINAL SUBJECT **AND** dependency(word2, word3) = NOMINAL SUBJECT **AND** dependency(word1, word4) = DIRECT OBJECT **AND** part-of-speech(word1) = VERB **AND** part-of-speech(word2) = VERB **AND** part-of-speech(word3) = NOUN **THEN** triples(word3, WORD2, word4)

3.3 Test Dataset

Table 5 shows the test dataset that was used in our experiment. It contains a list of 40 questions pertaining to various chapters in Al-Quran. It also includes the relevant documents where the answer for each question is located in Al-Quran, as in its chapter and verse i.e., "[Chapter:Verse]". The dataset has been adapted from the previous work [38], so that it caters to Muhammad Sarwar's version of English translation of Al-Quran. The questions remain the same but only the different words used in the questions are changed to the words that are used in Muhammad Sarwar's version of Quran translation.

Table 5: Test dataset

No.	Question	Relevant Verse
1	What did Imran's wife dedicate to her Lord?	**[3:35]** *Remember when Imran's wife prayed to her Lord saying, "I have made a vow to dedicate to Your service whatever is in my womb. Lord, accept it from me. You are All-hearing and All-knowing".*
2	What was Imran's baby named?	**[3:36]** *When the baby was born she said, "Lord, it is a female." God knew this. Male and female are not alike. "I have named her Mary. I pray that You will keep her and her offspring safe from Satan, the condemned one."* **[66:12]** *He has also told, as a parable, the story of Mary, daughter of Imran who preserved her virginity and (into whose womb) We breathed Our spirit. She made the words of her Lord and the predictions in His Books come true. She was an obedient woman.*

139

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

3	What did Imran's wife pray for her offspring?	**[3:36]** *When the baby was born she said, "Lord, it is a female." God knew this. Male and female are not alike. "I have named her Mary. I pray that You will keep her and her offspring safe from Satan, the condemned one."*
4	What did Zachariah find whenever he went to visit Mary in her place of worship?	**[3:37]** *Her Lord graciously accepted the offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her, "Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account.*
5	What happened when Zachariah was standing in his place of worship?	**[3:39]** *When he was standing during prayer in his place of worship, the angels called him saying, "God gives you the glad news of the birth of your son, John who will be a confirmation of (Jesus) the Word of God. He will become a chaste, noble leader and one of the righteous Prophets.*
6	What did Zachariah think was not possible since he is senile?	**[3:40]** *He said, "How can there be a son for me when I am already senile and my wife is barren." The angel replied, "God does as He wills."* **[19:8]** *He said, "Lord, how can I have a son? My wife is barren and I have reached an extremely old age".*
7	What did Abraham ask his father Azar?	**[6:74]** *Consider when Abraham asked his father, Azar, "Why do you believe idols to be your gods? I find you and your people in absolute error".*
8	What is salsabil?	**[76:18]** *from a spring named salsabil.*
9	What was revealed regarding the party of jinn?	**[72:1]** *(Muhammad), say, "It has been revealed to me that a party of jinn has listened (to the recitation) of the Quran and has told (their people), "We heard an amazing reading*
10	What is the story of the people of Sheba?	**[34:15]** *There was evidence (of the truth) for the people of Sheba in their homeland. (We gave them) two gardens, one on the left and one on the right and (told them), "Consume the sustenance which your Lord has given to you and give Him thanks. You have a blessed land and an All-forgiving Lord".*
11	What was the Prophet commanded to tell his wives?	**[33:28]** *Prophet, tell your wives, "If you want the worldly life and its beauty, I shall allow you to enjoy it and set you free in an honorable manner,* **[33:59]** *Prophet, tell your wives, daughters, and the wives of the believers to cover their bosoms and breasts. This will make them distinguishable from others and protect them from being annoyed. God is All-forgiving and All-merciful.*
12	What was the dream of Joseph?	**[12:4]** *When Joseph said, "Father, in my dream I saw eleven stars, the sun and the moon prostrating before me,"*
13	What happened in the city of Babylon?	**[2:102]** *They followed the incantations that the devils used against the kingdom of Solomon. Solomon did not hide the truth but the devils did. They taught magic to the people and whatever was revealed to the two angels, Harut and Marut, in Babylon. The two angels did not teach anything to anyone without saying, "Our case is a temptation for the people, so do not hide the truth." People learned something from the two angels that could cause discord between a man and his wife. However, they could harm no one except by the permission of God. In fact, the (people) learned things that would harm them and render them no benefit. They knew very well that one who engaged in witchcraft would have no reward in the life hereafter. Would that they had known that they had sold their souls for that which is vile!*
14	Should we accept our fathers as guardians if they prefer disbelief?	**[9:23]** *Believers, do not accept your fathers and brothers as your guardians if they prefer disbelief to faith, lest you be unjust.*
15	Whose custody was Mary in?	**[3:37]** *Her Lord graciously accepted the offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her, "Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account.*
16	Who gives sustenance without keeping an account?	**[2:212]** *The worldly life is made to seem attractive to the disbelievers who scoff at the faithful, but the pious, in the life hereafter, will have a position far above them. God grants sustenance (without account) to anyone He wants.* **[3:37]** *Her Lord graciously accepted the offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her,*

140

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

		"Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account. **[24:38]** *(They worship Him) so that God will reward their best deeds and give them more through His favors. God gives sustenance to whomever He wants without account.*
17	Who did Zachariah pray to for his offspring?	**[3:38]** *Zachariah prayed to his Lord there, saying, "Lord, grant me, by Your Grace, virtuous offspring. You hear all prayers".*
18	Who was Zachariah asked to commemorate in the morning and evening?	**[3:41]** *Zachariah prayed to God saying, "Lord, show me the evidence (that it is Divine revelation)." The Lord replied, "You must not speak to people for three days except with gestures. Commemorate your Lord often and glorify Him in the early mornings and the evenings."*
19	Who was chosen over all women?	**[3:42]** *"Behold," the angels told Mary, "God had chosen you, purified you, and given you distinction over all women.*
20	Who was given glad news about Jesus?	**[3:45]** *"Behold," the angels told Mary, "God has given you the glad news of the coming birth of a son whom He calls His Word, whose name will be Messiah, Jesus, son of Mary, who will be a man of honor in this life and the life to come, and who will be one of the ones nearest to God.*
21	Who creates what He wants?	**[3:47]** *(Mary) said, "How can there be a son for me when no mortal has touched me?" The angel replied, "That is how God creates whatever He wants. When He decides to do something He just orders it to exist and it comes into existence."* **[24:45]** *God has created every living being from water: Some of them creep on their bellies; some walk on two feet and some of them walk on four legs. God creates whatever He wants. He has power over all things.* **[39:4]** *Had God wanted to have a son, He would have chosen one from His creatures according to His will. God is too Exalted to have a son. He is One and Almighty.* **[42:49]** *To God belongs the heavens and the earth. He created whatever He wanted. He grants males, female, or pairs of.*
22	Who should we worship as Jesus mentioned?	**[3:51]** *God is my Lord as well as yours. Worship Him for this is the right path."* **[19:36]** *(Baby Jesus said), "Worship God who is my Lord as well as yours. This is the straight path".*
23	Who prayed that no disbeliever should be on earth?	**[71:26]** *Noah said, "Lord, do not leave a single disbeliever on earth;*
24	Who sends the clouds and raises the winds?	**[30:48]** *It is God who sends the winds to raise the clouds. He spreads them in the sky as He wants, then He intensifies them, and then you can see the rain coming down from the cloud. When He sends it down upon whichever of His servants He wants, they rejoice* **[35:9]** *It is God who sends the winds to raise the clouds. We then drive them unto barren areas and revive the dead earth. (The Resurrection) will also be executed in the same way.*
25	Who can hope in an indestructible bargain?	**[35:29]** *Those who recite the Book of God, who are steadfast in prayer and, who spend out of what We have given them for the cause of God, both in public and in private, have hope in an indestructible bargain*
26	Who brings forth the living from the dead?	**[6:95]** *It is God who makes all kinds of seeds grow, brings forth the living from the dead, and the dead from the living. It is God who does such things, so how can you turn away from Him?* **[30:19]** *He brings forth the living from the dead, takes out the dead from the living, and revives the earth from its death. Thus, you will all be brought back to life again.*
27	Who was killing the sons in the land?	**[2:49]** *(Children of Israel, recall My favor) of saving you from the Pharaoh's people who afflicted you with the worst kind of cruelty, slaying your sons and sparing your women. Your suffering was indeed a great trial from your Lord.* **[14:6]** *Moses told his people, "Remember the favors that God granted you when He saved you from the people of the Pharaoh who had punished you in the worst manner by murdering your sons and keeping your women alive. It was a great trial for you from your Lord.*

141

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

		[28:4] *The Pharaoh dominated the land and divided its inhabitants into different groups, suppressing one group by killing their sons and keeping their women alive. He was certainly an evil-doer.*
28	Where was Mary's food coming from?	**[3:37]** *Her Lord graciously accepted the offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her, "Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account.*
29	When was Zachariah given glad news of John?	**[3:39]** *When he was standing during prayer in his place of worship, the angels called him saying, "God gives you the glad news of the birth of your son, John who will be a confirmation of (Jesus) the Word of God. He will become a chaste, noble leader and one of the righteous Prophets.*
30	Why couldn't Zachariah speak to people for three days?	**[3:41]** *Zachariah prayed to God saying, "Lord, show me the evidence (that it is Divine revelation)." The Lord replied, "You must not speak to people for three days except with gestures. Commemorate your Lord often and glorify Him in the early mornings and the evenings."* **[19:10]** *Zachariah asked, " Lord, show me evidence (if this is a heavenly news)." The Lord said, "The evidence for it is that you must not speak (to any mortal) though you are in good health for three nights (and days)."*
31	Why would Allah send a Book?	**[6:7]** *(Muhammad), had We sent you a Book on paper, the unbelievers would have touched it with their hands but would still have said, "It is no more than plain sorcery".*
32	Why does Allah veil hearts?	**[6:25]** *Some of them listen to you, but We have veiled their hearts so that they cannot understand and made them deaf. They disbelieve all the evidence (of Our existence) that they may have seen. They only come to you for the sake of argument and the disbelievers say that (whatever Muhammad says) is no more than ancient legends.* **[18:57]** *Who are more unjust than those who are reminded of the revelations of their Lord but have disregarded them and have forgotten their deeds? We have veiled their hearts and sealed their ears so that they cannot understand. Even if you call them to the right path, they will never accept guidance.*
33	Why did Allah afflict nations with distress and adversity?	**[6:42]** *We had sent (Our guidance) to the nations who lived before you and afflicted them with distress and adversity so that they might submit themselves (to God).*
34	How did Mary have a son?	**[3:47]** *(Mary) said, "How can there be a son for me when no mortal has touched me?" The angel replied, "That is how God creates whatever He wants. When He decides to do something He just orders it to exist and it comes into existence."*
35	How did Jesus make a bird out of clay?	**[3:49]** *He will be a Messenger of God to the Israelites to whom he will say, "I have brought you a miracle from your Lord. I can create for you something from clay in the form of a bird. When I blow into it, it will become a real bird, by the permission of God. I can heal the blind and the lepers and bring the dead back to life, by the permission of God. I can tell you about what you eat and what you store in your homes. This is a miracle for you if you want to have faith.* **[5:110]** *When God said, "Jesus, son of Mary, recall My favors to you and your mother. (Recall) how I supported you by the holy spirit, made you speak to people from your cradle and when you grew up, taught you the Book, gave you wisdom, the Torah, and the Gospel. (Recall) when, by My will, you made a sculpture of a bird out of clay, blew into it, and it turned into a real bird by My Will. (Recall) how, by My will, you healed the deaf, the lepers, and raised the dead. (Recall) when you came to the Israelites in the house with clear miracles and I saved you from their mischief, even though the disbelievers among them said, "This is obviously magic".*
36	How can one hope in an indestructible bargain?	**[35:29]** *Those who recite the Book of God, who are steadfast in prayer and, who spend out of what We have given them for the cause of God, both in public and in private, have hope in an indestructible bargain*
37	How does rain come from the clouds?	**[24:43]** *Do you not see that God moves the clouds gently, brings them together, piles them up, and then you can see the rain coming from them. He*

142

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

		sends down hailstones from the mountains in the sky. With them He strikes or protects from them whomever He wants. The lightening can almost take away the sight. **[30:48]** *It is God who sends the winds to raise the clouds. He spreads them in the sky as He wants, then He intensifies them, and then you can see the rain coming down from the cloud. When He sends it down upon whichever of His servants He wants, they rejoice*
38	Which Surah mentions the story of the man who came running from farthest part of the city?	**[28:20]** *A man came running from the farthest part of the city saying, "Moses, people are planning to kill you. I sincerely advise you to leave the city.* **[36:20]** *A man came running from the farthest part of the city saying, "My people, follow the Messengers.*
39	Is there a waiting period to observe after divorce if there is no consummation of the marriage?	**[33:49]** *Believers, if you marry believing women and then divorce them before the consummation of the marriage, they do not have to observe the waiting period. Give them their provisions and set them free in an honorable manner.*
40	Was there a killing of the sons in the land?	**[2:49]** *(Children of Israel, recall My favor) of saving you from the Pharaoh's people who afflicted you with the worst kind of cruelty, slaying your sons and sparing your women. Your suffering was indeed a great trial from your Lord.* **[14:6]** *Moses told his people, "Remember the favors that God granted you when He saved you from the people of the Pharaoh who had punished you in the worst manner by murdering your sons and keeping your women alive. It was a great trial for you from your Lord.* **[28:4]** *The Pharaoh dominated the land and divided its inhabitants into different groups, suppressing one group by killing their sons and keeping their women alive. He was certainly an evil-doer.*

3.4 Query Formulation

Questions in natural language form such as in Table 5 are individually entered to the system as a query input. The query is parsed using Stanford CoreNLP to get the word dependencies of the query sentence as well as other NLP metadata (e.g., lemma, Part-of-Speech tag). This information is used to implement Semantic Dependency Triple Rules to generate semantic graphs for the query. Thus far, the query undergoes similar process of generating semantic graph for the knowledge base construction. Rather than populating the query graph into the knowledge base, the query graph becomes a reference for graph matching process. The system will find existing Quran graphs in the knowledge base that match with the query graph, in term of its nodes and relationship. As an example, Question (1) went through the processes above and its semantic graphs were generated as shown in Fig. 4.
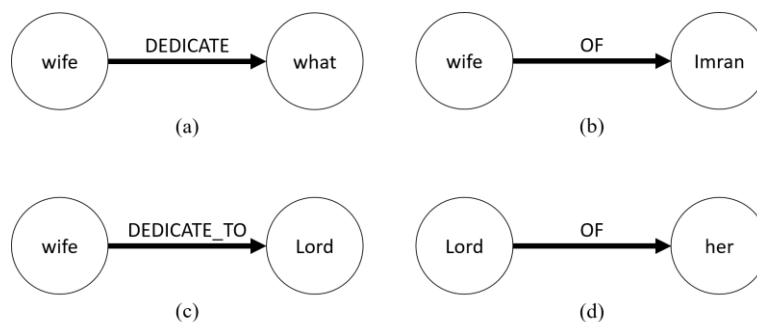

Fig. 4: Semantic graph generated for Question (1)

The figure shows four semantic graphs were generated for Question (1). Each of the query graphs are matched with the existing Quran graph in the knowledge base using Cypher query provided by Neo4j graph database platform. The exact matching nodes (keywords) and relationship (semantic relation) was searched. Quran graph that was successfully matched will also return information contained in its nodes and relationship, including its document (verse and chapter). Successful matching of multiple query graphs will result to a number of documents being collected. For example, query graph (a) might have been matched with Quran graph in several different documents, while graph (b) was matched with Quran graph in another set of documents. Based on this collection of documents, documents that have the highest number of occurrence are found and retrieved as the answer to the question. In another

143

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

words, a document or verse that has multiple graphs matching with the query graph indicates that it is highly relevant to the question. Therefore, the number of semantic graphs generated for the query can increase the accuracy of document retrieval of the system. This process was performed for the whole test dataset.

3.5 Evaluation Metrics

Precision and Recall are the standard evaluation metrics for knowledge retrieval systems. Precision is equivalent to the total number of relevant documents retrieved by the system over the total number of documents retrieved by the system [27], [44]. Recall is equivalent to the total number of relevant documents retrieved by the system over the total number of relevant documents for the query [27], [44]. F-score is also used as a measure for the accuracy of the proposed model [44]. The accuracy matrices for document retrieval are given in Eq. (1) – (3).

$$Precision = \frac{Total\ of\ relevant\ document\ retrieved}{Total\ of\ retrieved\ document} \tag{1}$$

$$Recall = \frac{Total\ of\ relevant\ document\ retrieved}{Total\ of\ relevant\ document} \tag{2}$$

$$F_{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

4.0 RESULT AND DISCUSSION

In this section, the experimental results and discussions are presented. Our knowledge representation model is designed for question answering including document retrieval applications. Since previous system [38] focused on retrieving documents relevant to the questions in their experiment, the proposed model is also tested on its document retrieval performance for the purpose of system comparison in this paper. An enhancement was made to the system and it was tested again. Both versions of our system was compared with previous research [38]. The previous system also focused on semantic knowledge representation of Al-Quran. The semantic representation was primarily based on noun and verb phrases. It used phrase structure parsing and logical formalism to extract noun and verb phrases in a sentence and they were stored in a graphical representation. Rather than phrase structure parsing, this paper proposes use of dependency parsing as it provides word dependencies which are comparable to the semantics of a sentence. The semantic dependency graph representation will be able to represent meanings of verses in Al-Quran better and perform well in document retrieval. Results for document retrievals of each system are shown in Table 6.

Table 6: Results for document retrieval of current and previous systems

No.	Relevant Document	Retrieved Document	
		Basic System	Enhanced System
1	[3:35]	[3:35], [3:37]	[3:35]
2	[3:36], [66:12]	NULL	[3:36]
3	[3:36]	[3:36]	[3:36]
4	[3:37]	[3:37]	[3:37]
5	[3:39]	[3:39]	[3:39], [3:37]
6	[3:40], [19:8]	NULL	[3:40]
7	[6:74]	[6:74]	[6:74]
8	[76:18]	NULL	NULL
9	[72:1]	[72:1]	[72:1]
10	[34:15]	[36:13], [9:70], [34:15]	[18:9], [36:13], [9:70], [34:15]
11	[33:28], [33:59]	[33:28], [33:59]	[33:28], [33:59]
12	[12:4]	[12:44]	[12:44], [12:4]
13	[2:102]	NULL	NULL
14	[9:23]	[2:145], [18:57], [9:23]	[9:23]
15	[3:37]	[3:37], [3:44]	[3:37], [3:44]

144

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

16	[2:212], [3:37], [24:38]	[24:38], [3:37]	[24:38], [3:37]
17	[3:38]	[3:38]	[3:38]
18	[3:41]	NULL	[3:41]
19	[3:42]	[3:42]	[3:42]
20	[3:45]	[19:97], [3:39], [3:45], [42:23], [9:21]	[19:97], [3:39], [3:45], [42:23], [9:21]
21	[3:47], [24:45], [39:4], [42:49]	[3:47], [42:49], [24:45]	[3:47], [42:49], [24:45]
22	[3:51], [19:36]	NULL	NULL
23	[71:26]	NULL	NULL
24	[30:48], [35:9]	NULL	NULL
25	[35:29]	[35:29], [6:95]	[35:29]
26	[6:95], [30:19]	[30:19]	[6:95], [30:19]
27	[2:49], [14:6], [28:4]	[28:4]	[28:4]
28	[3:37]	[3:37]	[3:37]
29	[3:39]	[19:97], [3:39], [3:45], [42:23], [9:21]	[19:97], [3:39], [3:45], [42:23], [9:21]
30	[3:41], [19:10]	[3:41]	[3:41], [19:10]
31	[6:7]	[2:213], [35:40], [18:1]	[6:7]
32	[6:25], [18:57]	NULL	NULL
33	[6:42]	[6:42]	[6:42]
34	[3:47]	NULL	[39:4]
35	[3:49], [5:110]	[5:110]	[5:110]
36	[35:29]	[35:29], [6:95]	[35:29]
37	[24:43], [30:48]	[24:43]	[24:43]
38	[28:20], [36:20]	[28:20], [36:20]	[28:20], [36:20]
39	[33:49]	[2:236], [2:237], [33:49]	[2:236], [2:237], [33:49]
40	[2:49], [14:6], [28:4]	NULL	NULL

The table shows experimental results for our basic system and the system with anaphoric pronouns enhancement. As can be seen in Table 6, the basic system managed to retrieve answers for 29 out of 40 questions asked. Other than Question (12) and (31), relevant document for all of the answered questions were successfully retrieved by the system. This shows that for the questions that it managed to answer, the system is able to retrieve relevant information to answer those questions. The system was able to retrieve 32 relevant documents out of the possible 60 relevant documents for the whole dataset.

Table 6 also shows that there were 11 questions that were not returned any document or returned "NULL". A number of reasons for such results were identified. Some of the unanswered questions have faced similar problems of anaphoric pronouns. Anaphoric pronouns problem can be described as inability to identify a set of entities which a grammatical unit (pronoun) is referring to [45]. Table 7 shows in details, several examples of document retrieval performed by both basic system as well as system with anaphoric pronouns enhancement.

Table 7: Examples of document retrieval for selected questions

Question	Retrieved Verse		Comment
	Basic System	Enhanced System	
(1) What did Imran's wife dedicate to her Lord?	**[3:35]** *Remember when Imran's wife prayed to her Lord saying, "I have made a vow to dedicate to Your service whatever is in my womb. Lord, accept it from me. You are All-hearing and All-knowing".* **[3:37]** *Her Lord graciously accepted the*	**[3:35]** *Remember when Imran's wife prayed to her Lord saying, "I have made a vow to dedicate to Your service whatever is in my womb. Lord, accept it from me. You are All-hearing and All-knowing".*	1. Query graphs generated are (wife, DEDICATE, what), (wife, DEDICATE_TO, Lord), (wife, OF, Imran) and (Lord, OF, she) using rules such as Direct Object Rule and Nominal Modifier with Verb Preposition Rule. 2. Both retrievals for basic system are the results of graph matching for (wife, OF, Imran) and (Lord, OF, she).

145

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

	offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her, "Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account.		3. Query graphs (wife, DEDICATE, what) and (wife, DEDICATE_TO, Lord) cannot be matched by basic system because the knowledge base does not recognize "I" and "wife" as the same entity. 4. The anaphoric pronouns enhancement to the basic system enable it to identify pronoun "I" in the verse "[3:35]" as referring to "wife". 5. The enhanced system is able to match all the query graphs and determine the most relevant verse is "[3:35]", ignoring irrelevant verses. The accuracy of the retrieval performance is improved.
(8) What is salsabil?	NULL	NULL	1. Query graph generated is (what, IS, salsabil) using Copular Rule. 2. The query graph finds no matching. The reason is information in the relevant verse "[76:18]" has not been represented in the knowledge base because of unidentified dependency obtained by Stanford CoreNLP tools i.e., "dep(spring, salsabil)". 3. The enhanced system faces the same problem of inaccurate dependency parsing.
(12) What was the dream of Joseph?	**[12:44]** *They replied, "It is a confused dream and we do not know the meaning of such dreams."*	**[12:44]** *They replied, "It is a confused dream and we do not know the meaning of such dreams."* **[12:4]** *When Joseph said, "Father, in my dream I saw eleven stars, the sun and the moon prostrating before me,"*	1. Query graph generated are (what, IS, dream) and (dream, OF, Joseph) using Copular Rule and Nominal Modifier with Of Preposition. 2. The retrieved verse "[12:44]" is irrelevant as it does not contain answer to the question. 3. The basic system fails to retrieve relevant verse "[12:4]", because the knowledge base stores the Quran graph (dream, OF, I). The pronoun "I" is not identified as "Joseph". Hence, the Quran graph is not matched. 4. The relevant verse "[12:4]" is retrieved after anaphoric pronoun enhancement is done to the system. The knowledge base identifies "I" in this verse as referring to "Joseph". The retrieval performance improves.
(16) Who gives sustenance without keeping an account?	**[24:38]** *(They worship Him) so that God will reward their best deeds and give them more*	**[24:38]** *(They worship Him) so that God will reward their best deeds and give them more*	1. Query graphs generated are (who, GIVE, sustenance) using Direct Object Rule.

146

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

	through His favors. God gives sustenance to whomever He wants without account. **[3:37]** *Her Lord graciously accepted the offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her, "Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account.*	*through His favors. God gives sustenance to whomever He wants without account.* **[3:37]** *Her Lord graciously accepted the offer and made Mary grow up, pure, and beautiful. Zachariah took custody of her. Whenever he went to visit her in her place of worship, he would find with her some food. He would ask her, "Where did this food come from?" She would reply, "God has sent it." God gives sustenance to whomever He wants without keeping an account.*	2. Graph matching results to retrieval of two relevant verses i.e., "[24:38]" and "[3:27]". 3. Another relevant verse "[2:212]" is not retrieved by the system because the semantic relationship used in the knowledge base is "grant" as in (God, GRANT, sustenance). The system is unable to identify "grant" and "give" as synonyms. 4. Results are the same for the enhanced system.
(18) Who was Zachariah asked to commemorate in the morning and evening?	NULL	**[3:41]** *Zachariah prayed to God saying, "Lord, show me the evidence (that it is Divine revelation)." The Lord replied, "You must not speak to people for three days except with gestures. Commemorate your Lord often and glorify Him in the early mornings and the evenings."*	1. Query graphs generated for are (Zachariah, COMMEMORATE_IN, morning) and (Zachariah, COMMEMORATE_IN, evening) using rules such as Nominal Modifier with Verb Preposition Rule. 2. The basic system is not able to match the query graph with Quran graphs in the knowledge base because it does not recognize pronoun "you" in Quran graph (you, COMMEMORATE_IN, morning) as the same entity as "Zachariah" in the query. This results to no matching graph. 3. The enhanced system is able to match the query graphs by identify pronoun "you" in the knowledge base as referring to "Zachariah". The retrieval performance is improved. 4. In term of question answering, pronoun "Him" in the knowledge base is also identified as "Lord", which allows the enhanced system to return the correct answer for this question.
(27) Who was killing the sons in the land?	**[28:4]** *The Pharaoh dominated the land and divided its inhabitants into different groups, suppressing one group*	**[28:4]** *The Pharaoh dominated the land and divided its inhabitants into different groups, suppressing one group*	1. Query graphs generated are (who, KILL, son) and (who, KILL_IN, land) using Direct Object Rule and Nominal

147

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

	by killing their sons and keeping their women alive. He was certainly an evil-doer.	*by killing their sons and keeping their women alive. He was certainly an evil-doer.*	Modifier with Verb Preposition Rule. 2. One of the relevant verses i.e., "[28:4]" is retrieved through graph matching for (who, KILL, son). 3. Other relevant verses are not retrieved as they use different words i.e., "slaying" and "murdering as opposed to "killing". Therefore, only verse that uses the same word as in the question is retrieved. 4. The enhanced system faces the previous problem of inability to identify synonyms of words used in the query.
(32) Why does Allah veil hearts?	NULL	NULL	1. No graph is generated for the query. The system is unable to apply appropriate rules because Stanford NLP tools inaccurately annotates "veil" as a noun instead of a verb, resulting to inaccurate set of dependencies including an unidentified dependency between "hearts" and "does" i.e., "dep(hearts, does)". 2. The enhanced system faces the previous problem of inaccurate dependency parsing.
(38) Which Surah mentions the story of the man who came running from farthest part of the city?	**[28:20]** *A man came running from the farthest part of the city saying, "Moses, people are planning to kill you. I sincerely advise you to leave the city.* **[36:20]** *A man came running from the farthest part of the city saying, "My people, follow the Messengers.*	**[28:20]** *A man came running from the farthest part of the city saying, "Moses, people are planning to kill you. I sincerely advise you to leave the city.* **[36:20]** *A man came running from the farthest part of the city saying, "My people, follow the Messengers.*	1. Query graphs generated are (Surah, MENTION, story), (story, OF, man), (man, CAME_FROM, part), (part, IS, farthest) and (part, OF, city) using rules such as Direct Object Rule, Nominal Modifier with Of Preposition Rule and Adjectival Modifier Rule. 2. All relevant verses are retrieved by matching of multiple graphs. 3. Results are the same for the enhanced system.

In Table 7, document retrievals by systems are shown for Questions (1), (8), (12), (16), (18), (27), (32) and (38). For example, information that is asked in Question (18) is regarding the entity "Zachariah", as specifically mentioned in the question. However, current knowledge base is storing the pronoun "you" in the relevant Quran graph i.e., (you, COMMEMORATE_IN, morning) and (you, COMMEMORATE_IN, evening), which were extracted from the original verse, "You must not speak to people for three days except with gestures. Commemorate your Lord often and glorify Him in the early mornings and the evenings". In this case, the system was not able to match "Zachariah" and "you" even though both were referring to the same entity. Similar problems were recurrent in Question (2), (6), (12), (26), (30), (31) and (34).

Other than anaphoric pronouns, the limitation of current Stanford dependency parsing has also contributed to the "NULL" results. Some of the questions were not correctly parsed by Stanford CoreNLP tools. For example, wrong dependencies were parsed by Stanford dependency parsing for Question (32). The word "veil" in the question was identified as a noun, which was not correct. The word "veil" in the sentence should be a verb instead. As a result of wrong set of dependencies provided, appropriate rules cannot be implemented and correct semantic graph cannot be

148

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

generated. On the other hand, unidentified dependency has been parsed for Questions (32) and (40). Unidentified dependency is labeled by Stanford CoreNLP when the dependency parser is not able to determine the precise relation between certain words. This can be due to complicated grammatical structure or conversion limitation of the parser itself. Further improvement on current dependency parser can ensure correct set of dependencies are provided to the system, resulting to proper graph matching process and accurate results.

Another reason for "NULL" returns is that the current graph matching mechanism is strictly focusing on matching the exact graph. Therefore, it is highly dependent on the capability of knowledge base construction module to generate all semantic graphs that can completely represent the whole Al-Quran. For instance, relevant documents for Question (8) and (13) can be retrieved through keyword matching. However, current system did not consider matching based on keywords only, to avoid huge amount of irrelevant documents being retrieved. This was because of the basic ranking and filtering mechanism implemented in current system. They can be improved in the future by including appropriate scoring mechanism to score keywords and even semantic relations.

In order to improve the retrieval performance of the current system, an enhancement was made to resolve one of the main problems, which was anaphoric pronouns. In this paper, anaphoric pronouns enhancement was performed to the basic system by correcting the anaphoric pronouns in the verses so that anaphoric reference between pronouns and entities is available. The co-reference features provided by Stanford CoreNLP tools were used to assist this process. The same experiment was repeated on the system with anaphoric pronouns enhancement. The complete experimental results are also shown in Table 6.

As can be seen in the table, the number of "NULL" results were reduced to 7 from 11 questions. The reasons for the persisting "NULL" results have been discussed earlier. The number of relevant documents that the system was able to retrieve also increased from 32 to 39 relevant documents for the whole dataset. With the help of anaphoric pronouns enhancement, the system now able to find more relevant Quran graphs to be matched with query graph, such as for Question (5), (12), (18), (26), (30) and (31). It helps to match semantics that are more appropriate to the question. More relevant documents were able to retrieved, while the number of irrelevant documents retrieved were reduced, as indicated by results for Question (1), (14), (25) and (36). This improved the overall accuracy of the system's retrieval. Other than anaphoric pronouns, synonym features can also be integrated to current system in the future. Lexical database such as WordNet can be integrated. This will improve the ability of the system in graph matching and retrieval. For instance, current system can only retrieve one out of three relevant documents for Question (27) because the different words are used in different verses which are "slaying", "murdering" and "killing" in "[2:49]", [14:6]" and "24:8", respectively.

Table 8: Document retrieval comparison for current and previous systems

	Wani's System	Basic System	Enhanced System
Total Document Retrieved	48	32	39
Total Relevant Document Retrieved	161	58	65

Based on Table 8, the number of relevant document that were retrieved by current system were 32 and 39 for basic system and enhanced system, respectively. Meanwhile the number of relevant document that were retrieved by previous system was 48 relevant documents. This suggests that the current system was not able to retrieve as many of the relevant document as the previous system. However, when we consider the total number of documents retrieved by the system, the current system improves in term of overall retrieval accuracy. The total number of documents retrieved by current system were only 58 and 65 for basic system and enhanced system, respectively. On the other hand, the total number of documents retrieved by the previous system was 161 documents. It indicates that although the previous system managed to retrieved many relevant documents, it was also retrieving many of the irrelevant documents. This has been improved by restricting the number of retrieval through precise matching of semantic graphs in the current system. Table 9 below shows further comparison between current systems (basic and enhanced systems) and previous system when evaluated using standard Precision, Recall and F-score metrics.

Table 9: Evaluation comparison for current and previous systems

Measurement	Wani's System	Basic System	Enhanced System
Precision	29.8%	53.3%	65.0%
Recall	80.0%	55.2%	60.0%
F-score	43.4%	54.2%	62.4%

149

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

As shown in Table 9, although the previous system scored high in Recall at 80% as it managed to retrieve many of the relevant documents, it scored poorly in Precision at 29.8% because of many irrelevant documents were also retrieved. The resulting F-score is also low at 43.4%. Current system has recorded higher Precision for both basic and enhanced systems at 53.3% and 65.0%, respectively. The Recall score for current systems is at middle range of 55.2% and 60% for basic and enhanced systems, respectively. As a result, the F-score for both of the current systems were also higher than the previous system. The basic system recorded F-score of 54.2% while the system with anaphoric pronouns enhancement recorded F-score of 62.4%. This indicates that the overall accuracy of document retrieval performance has been improved. The current system has achieved an increase of 108.1% for Precision and an increase of 43.8% for F-score. The nature of the proposed model that are designed for both question answering and document retrieval makes the retrieval to be more precise, in order to obtain the correct answer. Hence, the document retrieval becomes precise as well.

In term of improvement made by anaphoric pronouns enhancement, the system has improved its Precision and Recall by 21.5% and 8.7%, respectively. The overall accuracy also increased by 15.1%. It shows that the enhancement made to the system can improve current semantic knowledge representation. The results for the document retrieval experiments show that current system managed to produce acceptable results. For majority of the questions that were answered by the system, it successfully retrieves at least one relevant document. Irrelevant retrievals were also minimal. This allows users to spend less effort to go through the list of documents until the relevant document is found. This improvement is attributed to use of word dependencies in our Semantic Dependency Triple Rules. Some of the unanswered questions can be resolved by using better dependency parser, as current parser is still not capable in providing accurate and complete word dependencies for Quranic verses. Query expansion can also contribute to fewer unanswered questions by integration of lexical database or ontologies of world knowledge.

Current Semantic Dependency Triple Rules mainly covers simple rules such as Direct Object Rule and Nominal Modifier Rule, which are very useful to represent simple semantics. They are highly accurate and hardly misinterpreted. Other more complex rules can represent complicated semantics, but it depends on the ability of the parser to provide the accurate dependencies. As Semantic Dependency Triple Rules are developed further, and other enhancement are made, the scores for both Precision and Recall will become much higher.

5.0 CONCLUSION

This paper has proposed a semantic-graph knowledge representation model for Al-Quran based on word dependencies. The dependency features have been used in the proposed model to improve the accuracy for Al-Quran search applications that allow semantic-based search. The model is one of the formalisms for generating a semantic graph representation of Al-Quran. A set of rules called Semantic Dependency Triple Rules were defined to represents the logic of the meanings of the verse. The rules translated word dependencies and other metadata (such as part-of-speech and lemma) into a triple. The proposed model has been tested with Al-Quran on a document retrieval prototype and evaluated the accuracy of the retrieved document. The results show that the proposed model can be used to retrieve relevant documents with improved precision, compared to previous closely related system. Based on the experiment conducted, the proposed model can represent semantic knowledge of Quranic verses with satisfactory results.

The proposed model which focuses on representing semantics of Al-Quran results to more precise information storage. The rule-based formalism provides automatic knowledge base construction which costs less effort and time. The semantic graphical representation also is more expressive and flexible. The experiments conducted in this paper serve as indication to the capability of the proposed semantic graph representation in representing meaningful information in Al-Quran. The word dependencies are shown to able to represent semantics better than phrase structure in the sense that only meaningful information are extracted while irrelevant information are disregarded. In future research, the proposed model will be tested in a question answering system to study its capability in returning exact answer to a question from the relevant document. Further enhancement in term of rules refinement and integration with existing database can significantly improve the performance of current system in document retrieval.

6.0 ACKNOWLEDGEMENT

150

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

REFERENCES

[1] S. Saad, N. Salim, H. Zainal, and S. A. M. Noah, "A framework for Islamic knowledge via ontology representation," in *Proceedings - 2010 International Conference on Information Retrieval and Knowledge Management: Exploring the Invisible World, CAMP'10*, 2010, doi: 10.1109/INFRKM.2010.5466897.

[2] JavaTpoint, "Techniques of knowledge representation," *JavaTpoint*. [Online]. Available: https://www.javatpoint.com/ai-techniques-of-knowledge-representation. [Accessed: 24-Sep-2020].

[3] M. Singh, G. Kangri Vishwavidyalaya, P. Tanwar, T. V Prasad, M. S. Aswal, and A. Professor, "Comparative Study of Three Declarative Knowledge Representation Techniques," *Artic. Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 2, no. 7, pp. 2274–2281, 2010.

[4] H. T. Nguyen, V. H. Nguyen, and V. A. Vu, "A knowledge representation for Vietnamese legal document system," in *Proceedings - 2017 9th International Conference on Knowledge and Systems Engineering, KSE 2017*, 2017, pp. 30–35, doi: 10.1109/KSE.2017.8119430.

[5] F. Beirade, H. Azzoune, and D. E. Zegour, "Semantic query for Quranic ontology," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 6, pp. 754–760, 2021, doi: 10.1016/j.jksuci.2019.04.005.

[6] F. Harrag, A. Al-Nasser, A. Al-Musnad, R. Al-Shaya, and A. S. Al-Salman, "Quran Intelligent Ontology Construction Approach Using Association Rules Mining," *arXiv*, 2020.

[7] A. Ta'a, Q. A. Abed, B. M. Ali, and M. Ahmad, "Ontology-Based Approach for Knowledge Retrieval in Al-Quran Holy Book," *Int. J. Comput. Eng. Res. Ontol.*, vol. 06, no. 03, pp. 8–15, 2016.

[8] A. Hakkoum and S. Raghay, "Ontological approach for semantic modeling and querying the Qur'an," *Int. J. Islam. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 37–45, 2016.

[9] A. Hakkoum and S. Raghay, "Semantic Q&A System on the Qur'an," *Arab. J. Sci. Eng.*, vol. 41, no. 12, pp. 5205–5214, 2016, doi: 10.1007/s13369-016-2251-y.

[10] S. Wani, T. M. T. Sembok, and M. R. Wahiddin, "Constructing a knowledge base for Al-Qur'an utilizing principles of human communication," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2018, no. March, pp. 234–238, doi: 10.1109/INFRKM.2018.8464823.

[11] M. A. Mohd Yunus, A. Mustapha, R. Iqbal, and N. A. Samsudin, "An Ontological Approach towards Dialoguebased Information Visualization System: Quran Corpus for Juz' Amma," *MATEC Web Conf.*, vol. 135, pp. 1–9, 2017, doi: 10.1051/matecconf/201713500070.

[12] T. Weaam and S. Saad, "Ontology learning and population from quranic translation texts," in *International Conference on Advances in Information Technology for the Holy Quran and Its Sciences*, 2013, pp. 407–415.

[13] A. S. M. Rusli, F. Ridzuan, Z. M. Zaki, M. N. S. M. Sayuti, and R. A. Salam, "A systematic review on semantic-based ontology for Quranic knowledge," *Int. J. Eng. Technol.*, 2018, doi: 10.14419/ijet.v7i4.15.21376.

[14] S. Wani, T. M. T. Sembok, and M. R. Wahiddin, "Rule Based Modeling of Knowledge Bases," in *Proceedings - 2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017*, 2017, pp. 823–827, doi: 10.1109/CSCI.2017.142.

[15] S. Wani, T. M. T. Sembok, and M. S. Mir, "Context Aware Knowledge Bases for Efficient Contextual Retrieval: Design and Methodologies," in *Computational Science and Technology*, 2019, pp. 569–579.

[16] A. Taa, Q. A. Abed, and M. Ahmad, "Al-Quran ontology based on knowledge themes," *J. Fundam. Appl. Sci.*, vol. 9, no. 5S, pp. 800–817, 2017, doi: 10.4314/jfas.v9i5s.57.

[17]   T. R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," in *Formal Ontology in Conceptual Analysis and Knowledge Representation*, 1993, pp. 1–22.

[18]   A. T. McCray, "An upper-level ontology for the biomedical domain," *Comp. Funct. Genomics*, vol. 4, pp. 80–84, 2003, doi: 10.1002/cfg.255.

[19]   A. Ameen, K. U. R. Khan, and B. P. Rani, "Creation of ontology in education domain," in *Proceedings - 2012 IEEE 4th International Conference on Technology for Education, T4E 2012*, 2012, pp. 237–238, doi: 10.1109/T4E.2012.50.

[20]   M. M. Alqahtani and E. Atwell, "Developing Bilingual Arabic-English Ontologies of Al-Quran," in *2nd IEEE International Workshop on Arabic and Derived Script Analysis and Recognition, ASAR 2018*, 2018, pp. 96–101, doi: 10.1109/ASAR.2018.8480237.

[21]   H. Kharrazi and S. Raghay, "Collaborative Ontology Authoring in the Domain of the Holy Quran Knowledge," in *2019 7th Mediterranean Congress of Telecommunications (CMT)*, 2019, pp. 1–3.

[22]   A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad, "Ontology semantic approach to extraction of knowledge from Holy Quran," in *2013 5th International Conference on Computer Science and Information Technology, CSIT 2013 - Proceedings*, 2013, pp. 19–23, doi: 10.1109/CSIT.2013.6588752.

[23]   M. A. M. Safee *et al.*, "Hybrid search approach for retrieving Medical and Health Science knowledge from Quran," *Int. J. Eng. Technol.*, vol. 7, no. 4.15, pp. 69–74, 2018, doi: 10.14419/ijet.v7i4.15.21374.

[24]   R. H. Gusmita, Y. Durachman, S. Harun, A. F. Firmansyah, H. T. Sukmana, and A. Suhaimi, "A rule-based question answering system on relevant documents of Indonesian Quran Translation," in *2014 International Conference on Cyber and IT Service Management, CITSM 2014*, 2014, pp. 104–107, doi: 10.1109/CITSM.2014.7042185.

[25]   M. M. Alqahtani and E. Atwell, "Arabic quranic search tool based on ontology," in *International Conference on Applications of Natural Language to Information Systems*, 2016, pp. 478–485, doi: 10.1007/978-3-319-41754-7_52.

[26]   B. Hammo, H. Abu-Salem, S. L. Lytinen, and M. W. Evens, "QARAB: A Question Answering System to Support the Arabic Language," in *Proceedings of the ACL-02 Workshop on Computational Approaches to Semitic Languages*, 2002, p. 11.

[27]   H. Shmeisani, S. Tartir, A. Al-Na'ssaan, and M. Naji, "Semantically Answering Questions from the Holy Quran," in *International Conference on Islamic Applications in Computer Science And Technology*, 2014, pp. 1–8.

[28]   W. Alromima, R. Elgohary, I. F. Moawad, and M. Aref, "Applying ontological engineering approach for Arabic Quran corpus: A comprehensive survey," in *2015 IEEE 7th International Conference on Intelligent Computing and Information Systems, ICICIS 2015*, 2015, pp. 619–626, doi: 10.1109/IntelCIS.2015.7397287.

[29]   S. K. Hamed and M. J. A. Aziz, "A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification," *J. Comput. Sci.*, vol. 12, no. 3, pp. 169–177, 2016, doi: 10.3844/jcssp.2016.169.177.

[30]   N. Yusuf, M. A. M. Yunus, N. Wahid, N. M. Nawi, N. A. Samsudin, and N. Arbaiy, "Query expansion method for quran search using semantic search and lucene ranking," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 227–234, 2019.

[31]   A. Hakkoum and S. Raghay, "Advanced search in the Qur'an using semantic modeling," in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 2015, pp. 1–4, doi: 10.1109/AICCSA.2015.7507259.

152

Malaysian Journal of Computer Science. Information Retrieval and Knowledge Management Special Issue 2/2021

[32]   S. J. Putra, R. H. Gusmita, K. Hulliyah, and H. T. Sukmana, "A semantic-based question answering system for Indonesian translation of quran," in *ACM International Conference Proceeding Series*, 2016, pp. 504–507, doi: 10.1145/3011141.3011219.

[33]   R. A. Kadir, A. R. Yauri, and A. Azman, "SEMANTIC AMBIGUOUS QUERY FORMULATION USING STATISTICAL LINGUISTICS TECHNIQUE," *Malaysian J. Comput. Sci.*, pp. 48–56, 2018.

[34]   H. Abdelnasser *et al.*, "Al-Bayan : An Arabic Question Answering System for the Holy Quran," in *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, 2014, pp. 57–64.

[35]   F. S. Utomo, N. Suryana, and M. S. Azmi, "Question Answering Systems on Holy Quran: A Review of Existing Frameworks, Approaches, Algorithms and Research Issues," in *International Conference on Science & Technology (ICoST 2019)*, 2020, pp. 1–10, doi: 10.1088/1742-6596/1501/1/012022.

[36]   R. A. Kadir, A. R. Yauri, and A. Azman, "Automated Semantic Query Formulation for Document Retrieval," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2018, pp. 124–131, doi: 10.1109/INFRKM.2018.8464786.

[37]   S. Wani, M. R. Wahiddin, T. Mohd, and T. Sembok, "Logico-linguistic Semantic Representation of Documents," *2016 IEEE 14th Intl Conf Dependable, Auton. Secur. Comput. 14th Intl Conf Pervasive Intell. Comput. 2nd Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr.*, pp. 773–780, 2016, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.135.

[38]   S. Wani, "Al-Qur'an: Knowledge Representation and Question Answering," International Islamic University Malaysia, 2015.

[39]   M. De Marneffe and C. D. Manning, "The Stanford typed dependencies representation," in *Proceedings ofthe workshop on Cross-Framework and Cross-Domain Parser Evaluation*, 2008, pp. 1–8, doi: 10.3115/1608858.1608859.

[40]   M. De Marneffe and C. D. Manning, "Stanford typed dependencies manual," no. September. pp. 1–28, 2016.

[41]   C. D. Manning, J. Bauer, J. Finkel, and S. J. Bethard, "The Stanford CoreNLP Natural Language Processing Toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014, pp. 55–60.

[42]   R. A. Kadir and A. R. Yauri, "Resource description framework triples entity formations using statistical language model," *J. Fundam. Appl. Sci.*, vol. 9, no. 4S, p. 710, 2018, doi: 10.4314/jfas.v9i4s.40.

[43]   G. Schreiber and Y. Raimond, "Rdf 1.1 Primer," *W3C Working Group Note 24 June 2014*, 2014. [Online]. Available: https://www.w3.org/TR/rdf11-primer/.

[44]   A. M. N. Allam and M. H. Haggag, "The Question Answering Systems: A Survey," *Int. J. Res. Rev. Inf. Sci.*, vol. 2, no. 3, 2012.

[45]   C. L. Miller-Naudé and J. A. Naudé, "Theoretical Approaches to Anaphora and Pronouns in Biblical Hebrew," *J. Semit.*, vol. 28, no. 2, pp. 1–22, 2019, doi: 10.25159/2663-6573/7262.