SEMANTIC GRAPH KNOWLEDGE REPRESENTATION OF AL-QURAN FOR QUESTION ANSWERING SYSTEM

MUHAMMAD MUHTADI BIN MOHAMAD KHAZANI

MASTER OF SCIENCE
(COMPUTER SCIENCE)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

2022

SEMANTIC GRAPH KNOWLEDGE REPRESENTATION OF AL-QURAN FOR QUESTION ANSWERING SYSTEM

MUHAMMAD MUHTADI BIN MOHAMAD KHAZANI

Thesis submitted to Centre for Graduate Studies, Universiti Pertahanan Nasional Malaysia, in fulfilment of the requirements for the Degree of Master of Science (Computer Science)

2022

ABSTRACT

Al-Quran is one of the primary knowledge resources in Islam, containing a vast amount of knowledge in various domains. Most of the current knowledge representation models for Al-Quran are based on the ontological approach, which focuses on extracting concepts in Al-Quran rather than meanings of the Quranic verses. There is a lack of research that focuses on utilizing word dependencies for capturing the meanings of Quranic verses. This research proposes a semantic graph knowledge representation for Al-Quran using word dependencies. The proposed model obtains dependencies between the words in Quranic text through dependency parsing. Based on syntactic and semantic analyses, a set of rules is developed for generating semantic triples representing the meanings of the Quranic verses based on their word dependencies. The semantic triples are mapped into a graph database as semantic dependency graph. The knowledge representation model has been tested in a question answering experiment. The results are evaluated for retrieval accuracy using Precision, Recall and F-score metrics. The proposed model has achieved 62.7%, 53.3%% and 57.7% for the respective metrics. Therefore, the rule-based semantic graph representation model for Al-Quran is a viable approach to represent semantic knowledge of Quranic verses.

ABSTRAK

Al-Quran merupakan salah satu sumber ilmu yang utama dalam Islam, yang mengandungi sejumlah besar ilmu pengetahuan dalam pelbagai bidang. Kebanyakan model perwakilan pengetahuan semasa bagi Al-Quran adalah berasaskan pendekatan ontologi, yang menumpu kepada mengestrak konsep-konsep dalam Al-Quran berbanding makna setiap ayat. Terdapat kekurangan penyelidikan yang memberi fokus kepada kebersandaran perkataan bagi memetik makna setiap ayat Al-Quran. Kebanyakan sistem capaian pengetahuan Al-Quran sedia ada adalah berasaskan carian kata kunci, yang mungkin memberikan jawapan yang tidak relevan kepada pertanyaan. Penyelidikan ini mencadangkan sebuah perwakilan pengetahuan graf semantik bagi Al-Quran menggunakan kebersandaran perkataan. Model yang dicadangkan memperoleh kebersandaran antara perkataan-perkataan di dalam teks Al-Quran melalui penghurai kebersandaran. Berdasarkan analisis sintaks dan semantik, satu set petua dibangunkan bagi menghasilkan rangkap tiga semantik yang mewakili makna kepada ayat Al-Quran berasaskan kebersandaran perkataannya. Rangkap tiga semantik tersebut dipetakan ke dalam sebuah pangkalan data graf sebagai graf kebersandaran semantik. Model perwakilan pengetahuan tersebut telah diuji melalui eksperimen soal jawab. Hasil keputusan dinilai dengan ketepatan capaian menggunakan metrik Kejituan, Dapatan Semula dan Skor-F. Model yang dicadangkan telah mencapai skor masing-masing 62.7%, 53.3%% dan 57.5% bagi metrik tersebut. Dengan itu, model perwakilan graf semantik berasaskan peraturan bagi Al-Quran merupakan satu pendekatan yang berupaya untuk mewakilkan pengetahuan semantik bagi ayat-ayat Al-Quran.

ACKNOWLEDGEMENT

All praise be to Allah. Salutations and greetings to the Prophet Muhammad p.b.u.h. My utmost gratitude to the Almighty Allah for giving me the strength and motivation to complete this Master's degree.

'I would like to express my deep and sincere gratitude to my supervisor, Dr. Hassan bin Mohamed as the main supervisor who has continuously guided me to this research completion. His patience, guidance, and advice have helped me learn so much throughout this research. I would also like to extend my sincere thanks to Prof. Emeritus Dato' Dr. Tengku Mohd bin Tengku Sembok and Prof. Madya Dr. Nurhafizah Moziyana binti Mohd Yusof for their valuable advice and comments as co-supervisors.

My warmest thanks to Dr. Sharyar Wani and Dr. Syahaneim binti Marzukhi for their insightful comments and suggestions in the proposal defence and research progress colloquium sessions. A token of gratitude to the staff at Cyber Security and Digital Industrial Revolution Centre for the friendly and conducive working environment. Special thanks to the staff at the campus for their assistance. Thanks to the financial support from Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (FRGS/1/2019/ICT02/UPNM/01/1) that was channelled to the Universiti Pertahanan Nasional Malaysia (UPNM).

Last but not least, I am very grateful to my parents and family members for their unconditional support and words of encouragement. May Allah shower His blessing upon them.

APPROVAL

The Examination Committee has met on **14 September 2022** to conduct the final examination of **Muhammad Muhtadi bin Mohamad Khazani** on his/her master/PhD thesis entitled '**SEMANTIC GRAPH KNOWLEDGE REPRESENTATION OF AL-QURAN FOR QUESTION ANSWERING SYSTEM**'.

The committee recommends that the student be awarded the Master of Science (Computer Science)

Members of the Examination Committee were as follows

Prof. Madya Dr. Mohd Fahmi bin Mohamad Amran
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Chairman)

Prof. Madya Dr. Zuraini binti Zainol
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

Prof. Madya Dr. Nazlia binti Omar
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
(External Examiner)

APPROVAL

This thesis was submitted to the Senate of Universiti Pertahanan Nasional Malaysia and has been accepted as fulfilment of the requirements for the degree of **Master of Science (Computer Science)**. The members of the Supervisory Committee were as follows.

Dr. Hassan bin Mohamed
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Main Supervisor)

Prof. Emeritus Dato' Dr. Tengku Mohd bin Tengku Sembok
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Co Supervisor)

Prof. Madya Dr. Nurhafizah Moziyana binti Mohd Yusop
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Co Supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

DECLARATION OF THESIS

Student's full name : MUHAMMAD MUHTADI BIN MOHAMAD KHAZANI

Date of birth : 28 FEBRUARY 1992

Title : SEMANTIC GRAPH KNOWLEDGE REPRESENTATION OF AL-QURAN FOR QUESTION ANSWERING SYSTEM

Academic session : MARCH 2020 – JULY 2022

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

I further declare that this thesis is classified as:

☐ **CONFIDENTIAL** (Contains confidential information under the official Secret Act 1972)*

☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*

☐ **OPEN ACCESS** I agree that my thesis to be published as online open access (full text)

I acknowledge that Universiti Pertahanan Nasional Malaysia reserves the right as follows.

1. The thesis is the property of Universiti Pertahanan Nasional Malaysia.
2. The library of Universiti Pertahanan Nasional Malaysia has the right to make copies for the purpose of research only.
3. The library has the right to make copies of the thesis for academic exchange.

_____	_____
Signature	**Signature of Supervisor/Dean of CGS/ Chief Librarian
_____	_____
IC/Passport No.	**Name of Supervisor/Dean of CGS/ Chief Librarian
Date:	Date:

*If the thesis is CONFIDENTAL OR RESTRICTED, please attach the letter from the organisation with period and reasons for confidentiality and restriction.
** Witness

TABLE OF CONTENTS

TITLE		PAGE
ABSTRACT		**ii**
ABSTRAK		**iii**
ACKNOWLEDGEMENT		**iv**
APPROVAL		**v**
DECLARATION OF THESIS		**vii**
TABLE OF CONTENTS		**viii**
LIST OF TABLES		**x**
LIST OF FIGURES		**xiii**
LIST OF ABBREVIATIONS		**xiv**
CHAPTER 1	**INTRODUCTION**	**1**
	1.1 Research background	1
	1.1.1 Al-Quran knowledge representation	5
	1.1.2 Al-Quran knowledge retrieval	6
	1.2 Problem statement	7
	1.3 Research objectives	9
	1.4 Research questions	10
	1.5 Research scope and limitations	10
	1.6 Significance of the research	11
	1.7 Thesis Outline	12
CHAPTER 2	**LITERATURE REVIEW**	**14**
	2.1 Introduction	14
	2.2 Knowledge representation	15
	2.3 Knowledge representation models for Al-Quran	16
	2.3.1 The Holy Quran	16
	2.3.2 Semantic-based knowledge representation models for Al-Quran	18
	2.3.3 Rule-based approaches	29
	2.3.4 Graph-based approaches	31
	2.4 Question answering system	34
	2.4.1 Keyword-based and semantic-based search	36
	2.5 Conclusion	40
CHAPTER 3	**METHODOLOGY**	**41**
	3.1 Introduction	41
	3.2 Conceptual framework	41
	3.2.1 Knowledge base construction	43
	3.2.2 Question answering	46
	3.3 Corpus and data collection	48
	3.3.1 Quran corpus	48
	3.3.2 Dataset	50
	3.4 Syntactic and semantic analysis	57

3.4.1 Natural language processing 57
3.4.2 Semantic pattern analysis 59
3.5 Rules Formulation 63
3.5.1 Semantic triples 64
3.5.2 Semantic graph 78
3.6 Proof of concept 83
3.6.1 Graph database 83
3.6.2 Question answering prototype 85
3.7 Evaluation 93
3.7.1 Dataset 93
3.7.2 Metric 94
3.8 Conclusion 94

CHAPTER 4 RESULTS AND DISCUSSIONS 96
4.1 Introduction 96
4.2 Result and analysis 96
4.2.1 Proposed system 97
4.2.2 Edited knowledge base with anaphoric pronoun resolution 104
4.2.3 Comparative analysis 111
4.3 Discussions 114
4.3.1 Semantic graph representation for Al-Quran 114
4.3.2 Availability of NLP tools for Quranic text 115
4.3.3 Integration of additional features, databases or ontologies 115
4.3.4 Performance in question answering 116
4.4 Conclusion 116

CHAPTER 5 CONCLUSION 118
5.1 Introduction 118
5.2 Contribution 119
5.3 Future work 123
5.4 Conclusion 124
REFERENCES 126
APPENDICES 134
Appendix A 134
Appendix B 135
Appendix C 137
Appendix D 140
Appendix E 143
BIODATA OF THE STUDENT 147
LIST OF PUBLICATIONS 148

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 3.1	Comparison number of unidentified dependencies of parsing result on selected chapter from various translation of Al-Quran	50
Table 3.2	Original dataset (Wani, 2016)	50
Table 3.3	Comparison between questions in altered dataset and original dataset	52
Table 3.4	Altered dataset for Quran question answering	52
Table 3.5	Example of suggested meanings and triples for "I have made a vow to Your service whatever is in my womb."	59
Table 3.6	Suggested meanings and triples for a sentence of Verse 35 of Surah Ali Imran “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.”	63
Table 3.7	Sentences in Verse 35, Surah Ali 'Imran	65
Table 3.8	Suggested meanings and triples for the first sentence of Verse 35 of Surah Ali ‘Imran “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.”	65
Table 3.9	Suggested meanings and triples for the second sentence of Verse 35 of Surah Ali ‘Imran “Lord, accept it from me."	65
Table 3.10	Suggested meanings and triples for the third sentence of Verse 35 of Surah Ali ‘Imran “You are All-hearing and All-knowing"	66
Table 3.11	Dependencies for the sentence "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." of Verse 35, Surah Ali 'Imran	68
Table 3.12	Dependencies for the sentence "Lord, accept it from me." of the Verse 35, Surah Ali 'Imran	69
Table 3.13	Dependencies for the sentence "You are All-hearing and All-knowing”." of the Verse 35, Surah Ali 'Imran	69
Table 3.14	NLP metadata for the sentence "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." of Verse 35, Surah Ali 'Imran	70
Table 3.15	NLP metadata for the sentence "Lord, accept it from me." of Verse 35, Surah Ali 'Imran	70
Table 3.16	NLP metadata for the sentence "You are All-hearing and All-knowing”." of the Verse 35, Surah Ali 'Imran	71
Table 3.17	Examples of semantic triple of a subject-verb-object pattern and their word dependencies and Part-Of-Speech tags	72
Table 3.18	Direct Object Rule	73
Table 3.19	Application of Rule 1 towards Verse 35 of Surah Ali ‘Imran	74
Table 3.20	Nominal Modifier With Place Preposition	74
Table 3.21	List of Semantic Dependency-Triple Rules	75
Table 3.22	Semantic Triples generated based on Semantic Dependency-Triple Rules from Verse 35 of Surah Ali 'Imran	77
Table 3.23	Semantic Triple-Graph Rules	80
Table 3.24	Example of semantic triples converted to semantic graphs	81
Table 3.25	Example of application of rules on query “What did Imran's wife dedicate to her Lord?”	87
Table 3.26	Semantic triples generated for query “What did Imran's wife dedicate to her Lord?”	87
Table 3.27	Example of Cypher command for graph matching	91
Table 3.28	Example of graph matching returns for query "What did Imran’s wife pray for her offspring?"	92
Table 3.29	Example of document ranking for query "What did Imran’s wife pray for her offspring?"	92
Table 4.1	Experimental results for proposed system on document retrieval	97
Table 4.2	Example of retrievals by proposed system	99
Table 4.3	Example of irrelevant retrievals caused by anaphoric pronouns problem	101
Table 4.4	Example of “NULL” retrievals caused by parsing error	102

Table 4.5	Examples of "NULL" retrievals caused by unavailable graph in the knowledge base	103
Table 4.6	Experimental results for edited knowledge base on document retrieval	105
Table 4.7	Example of increased graph matching by the edited knowledge base	107
Table 4.8	Example of reduced number of irrelevant retrievals by the edited knowledge base	109
Table 4.9	Example of limited retrievals caused by different keywords	111
Table 4.10	Document retrieval comparison for proposed system and previous system (Wani, 2016)	112
Table 4.11	Evaluation comparison for proposed system and previous system (Wani, 2016)	113
Table 6.1	Part-Of-Speech tagset used in Semantic Dependency Graph Rules (Marcus et al., 1993)	134
Table 6.2	List of grammatical relations	135

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 3.1	Conceptual framework for knowledge base construction and question answering	42
Figure 3.2	Dependencies of a sentence "John is going to Boston by bus."	44
Figure 3.3	Dependency parse tree of a sentence "John is going to Boston by bus."	45
Figure 3.4	Text processing of Quranic text	58
Figure 3.5	Dependency parse tree for the first sentence of Verse 25 of Surah Ali 'Imran	61
Figure 3.6	Dependency parse tree obtained for the second sentence in Verse 35 of Surah 'Ali Imran “Lord, accept it from me.”	67
Figure 3.7	Dependency parse tree obtained for the third and last sentence in Verse 35 of Surah 'Ali Imran “You are All-hearing and All-knowing”	67
Figure 3.8	Direct object pattern	73
Figure 3.9	A graphical graph representation of a semantic triple	79
Figure 3.10	Semantic graph representation of Verse 35 of Surah Ali ‘Imran in Neo4j graph database	82
Figure 3.11	Semantic graph representation for Al-Quran in Neo4j graph database	83
Figure 3.12	Example of graph matching for all components	89
Figure 3.13	Example of graph matching with partial matching on relationship only	90
Figure 6.1	Semantic Dependency Graph for Surah Ali ‘Imran, Chapter 3, Verse 36	137
Figure 6.2	Semantic Dependency Graph for Surah Ali ‘Imran, Chapter 3, Verse 39	138
Figure 6.3	Semantic Dependency Graph for Surah At-Taubah, Chapter 9, Verse 23	139

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
API	-	Application Programming Interface
ATE	-	Attributional Triple Encoder
BAMA	-	Buckwalter Arabic Morphological Analyzer
CRUD	-	Create, Read, Update, and Delete
DL	-	Description Logic
GOLD	-	General Ontology for Linguistic Description
HNC	-	Hierarchical Network of Concepts
HQA	-	Hybrid Question Answering
ICTCLAS	-	Institute of Computing Technology, Chinese Lexical Analysis System
IRQA	-	Information Retrieval Question Answering
KBQA	-	Knowledge Base Question Answering
KR-EAR	-	Knowledge Representation Model with Entities, Attributes and Relations
MAP	-	Mean Average Precision
MOHE	-	Ministry of Higher Education
MOSTI	-	Ministry of Science, Technology and Innovation
MRR	-	Mean Reciprocal Rank
NLP	-	Natural Language Processing
NLPQA	-	Natural Language Processing Question Answering
NN	-	Neural Network
ORM	-	Object-Role Modelling
OWL	-	Web Ontology Language
QAC	-	Quran Arabic Corpus
QCO	-	Quranic Corpus Ontology
RDF	-	Resource Description Framework
REST	-	Representational State Transfer
RTE	-	Relational Triple Encoder
SMO	-	Semantic Miracle Ontology
SPARQL	-	Simple Protocol and RDF Query Language

SVM	-	Support Vector Machine
TF-IDF	-	Term Frequency Inverse Document Frequency
UPNM	-	Universiti Pertahanan Nasional Malaysia
USIM	-	Universiti Sains Islam Malaysia
WEKA	-	Waikato Environment for Knowledge Analysis
YAGO	-	Yet Another Great Ontology

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Part-Of-Speech tagset	134
Appendix B	List of grammatical relations	135
Appendix C	Examples of Semantic Dependency Graph of Quranic verses	137
Appendix D	Examples of question answering of query	140
Appendix E	Number of matching graph for document retrieved by the proposed system	143

CHAPTER 1

INTRODUCTION

1.1 Research background

This study concerns with semantic representation of knowledge in Al-Quran. The ultimate goal of knowledge representation studies is to enable computer to attain the basic "common senses" of the world and subsequently generate new information by applying and adapting to new circumstances. Therefore, studies have focused on solving how to impart this common sense to computer. Levesque (1986) states that despite the apparent simplicity of the said goal, there are various ideas regarding the fundamental issues in knowledge representation. A representation and classification approach which can be understood by machines and humans is necessary to help users recognize the precise semantics of knowledge in complex text (Ta'a et al., 2017).

A large volume of documents has been produced in the natural language form. Various knowledge representation languages such as logic, semantic networks, frame language and production rules have been used to extract knowledge contained in the text documents. Syntax-based approaches for knowledge representation were able to extract a sufficient amount of information from the text. Additionally, semantic-based

techniques could enrich the knowledge representation and improve retrieval performance of the knowledge base. A set of criteria was proposed by Katalnikova & Novickis (2018) for knowledge representation models. The set of criteria defines a number of requirement clusters that should be fulfilled by a good knowledge representation model such as knowledge meaning representation, knowledge representation in natural language notions, knowledge hierarchical structure representation etc. These criteria were collected through analyzing various requirements that have been submitted to a knowledge representation model. The semantic network model was regarded as the closest representation of knowledge in natural language texts because its properties align with the proposed criteria.

Hierarchical rules were applied by Sakharov (2019) in his predicate representations learning framework. The validity probability of First Order Logic formulas were computed using the predicate representations learned through Neural Network (NN) training. These probabilities are useful for building knowledge bases and making goal-directed decisions. The hierarchical rules applied in this study are an extension of Horn clauses. The hierarchical rules provide positive and negative facts which indicate the polarity of a literal (an atomic formula). A positive fact is basically a literal, while negative fact is the negation of a literal. This is necessary for the NN learning in their research. The set of facts and rules define three kinds of predicates in the knowedge base which are extensional, decidable and hierarchically definable. Hierarchical rules allow for one-by-one learning of the predicates, resulting in a shorter learning time and higher accuracy of learned predicate representations.

A framework for knowledge extraction using frames was proposed by Corcoglioniti et al. (2016). It identified semantic frames in a text and extract entities and relationships between them. The proposed framwork was designed with two phases for the knowledge extraction. The linguistic feature extraction phase involved natural language processing tasks to build an Resource Description Framework (RDF) graph of mentions. Whilst the knowledge distillation phase applied an extended RDF Processor tool and Simple Protocol and RDF Query Language (SPARQL)-like mapping tools to manipulate the graph of mentions and represent knowledge imparted in a text as a knowledge graph.

A hybrid approach combining frames language and production rules was presented by Nguyen et al. (2017) to extract knowledge from legal documents. The frames language was applied in the definition of entities, categories, and relationships of the documents. The production rules were used in their reasoning algorithm for knowledge structure manipulation. The proposed approach is capable of automatically locating authentic legal documents with respect to the timeline. Frame-based knowledge representations can be applied in various fields including health care, forecasting as well as natural language processing (Nazaruks & Osis, 2017). Existing frame-based knowledge systems are often integrated with ontology nets.

A new Knowledge Representation Model with Entities, Attributes and Relations (KR-EAR) was presented by Lin et al. (2016), improving on the previous Translation in Embedding Space (TransE), Translation on Hyperplanes (TransH) and Translation in Relation Space (TransR) knowledge representation models. Relations in a knowledge graph were divided into attributes and relations using Relational Triple

Encoder (RTE) and Attributional Triple Encoder (ATE), so that the knowledge representation model can learn more reasonable embeddings for entities and relations.

A multidimensional structure of the knowledge base was also introduced to help resolve one of the main challenges of knowledge discovery which is the extraction of meaningful information from data in context (Zenkert & Fathi, 2016). They have used text mining, analytics and multidimensional knowledge base construction to achieve better performance in knowledge extraction.

Wani et al. (2016) presented a hybrid approach for knowledge extraction and representation that combines logic and linguistics. It helps computer to perform data analysis efficiently and produce accurate retrievals. The logico-linguistic strategy considerably improved the precision of natural language knowledge translation.

Different knowledge representation languages have been used for various strategies which ultimately share the same goal to produce a comprehensive represention of meaningful knowledge. Through these knowledge representation methods, researchers were able to extract meaningful knowledge from textual documents and utilize them in a variety of applications such as information retrieval and question answering.

1.1.1 Al-Quran knowledge representation

The Holy Al-Quran is one of the most influential text documents in the world. It is regarded as one of the primary resources for the Islamic knowledge as well as the Arabic language (Saad et al., 2010). Al-Quran is a unique and complex corpus which contains vast knowledge to be explored. Thus, knowledge that resides within the Holy Quran should be represented in the most appropriate manner and efficient mechanism for further knowledge processing and inferencing. Most of the knowledge representation models for Al-Quran are based on ontology. Ontology is a formal, explicit specification of a shared conceptualization (Gruber, 1995). It represents a collection of facts and concepts of a particular domain and its relation. Additionally, it can be seen that most of the existing knowledge representation models focused on Arabic Al-Quran and English translations.

In a Quran ontology by Harrag et al. (2020), concepts and semantic relations from Al-Quran were extracted using linguistic pattern-based schemes and associations rules. They have applied the Apriori algorithm to generate a set of rules for concept extraction. The ontology is said to be able to provide a more expressive representation of Quran relations in terms of rules.

An application of logic and linguistic approach in the design and development of a comprehensive knowledge base for Al-Quran was proposed by Wani et al. (2018b). Noun-verb structures linked to the event description, events and role analysis become the primary focus of syntactic analysis of the text and are translated to predicate-subject forms. On top of that, semantic relatedness and association between

concepts are preserved using semantic graph. Graphical approaches are generally popular for knowledge representation and natural language semantics (Fernandes & Bernardino, 2018). Therefore, this approach can be used to represent semantics of Quranic text.

1.1.2 Al-Quran knowledge retrieval

The goal of knowledge retrieval from a large volume of information is to provide user support for interpreting and obtaining adequate information (Ta'a et al., 2017). In the initial stage of the knowledge representation research, much research was carried out using keyword-based matching (Yauri et al., 2013). This method basically starts by finding possible combinations of keywords that are relevant to the desired words. Currently, researchers use semantic-based method which helps to retrieve the information through similarity techniques to match the contextual meaning of keyword. Most of the knowledge retrieval systems are used in information retrieval systems using keyword matching techniques. Query expansion techniques are also commonly used to improve accuracy rate.

A Quran application known as QARAB was developed by Hammo & Lytinen (2002), which was the result of combining traditional information retrieval techniques with a sophisticated Natural Language Processing (NLP) approach. The application was to identify text passages that answer a natural language question. Abdelnasser et al. (2014), on the other hand, developed Al-Bayan with aims of understanding the semantics of Al-Quran and answering users' questions using reliable Quranic

resources. They applied Support Vector Machine (SVM) to classify questions with ranked answers for high accuracy.

Utomo et al. (2020) have mentioned in their review that common question answering systems have used many techniques for answer retrievals such as keyword matching, keyword expansion, SPARQL query, SVM Question classification, cosine similarity, Artificial Neural Network (ANN) Verb classification and N-gram techniques in order to achieve better retrieval results. However, research on question answering systems for Al-Quran seems to be scarce and most of them are in Arabic. Most of the systems also focus on a restricted domain, and they are able to answer factoid or simple statement queries.

1.2 Problem statement

The problem of computer understanding natural language is a constant challenge in the artificial intelligence research area. Researchers have been taking various approaches to solve this complex problem, and it has always been the longstanding goal. (Stanojević & Vraneš, 2005) suggest that the natural language understanding problem can be deconstructed into three components: natural language grammar definition, meaning representation, and natural language input parse.

The traditional keyword-based approach to knowledge representation of Al-Quran does not preserve semantic relatedness between words in Quranic verses (Suganya et al., 2013). Keyword-based information retrieval systems also depend on exact keyword matching, which might cause irrelevant retrievals (Gusmita et al., 2014;

Hamed & Aziz, 2016; Shmeisani et al., 2014). Semantic approach has been used to improve knowledge representation as well as retrieval of Al-Quran, as it takes into account the meanings of words in a sentence.

Most of the existing knowledge representation models for Al-Quran are using ontological representation (Abbas, 2009; Moogab et al., 2021; Sadi et al., 2016; Safee et al., 2018; Sherif & Ngonga Ngomo, 2015; Ullah Khan et al., 2013). Ontology provides certain advantages for knowledge base application, such as knowledge sharing and reusability. However, ontology construction is time-consuming and requires a lot of work. There needs to be effortless and efficient knowledge representation model for Al-Quran.

Current Quran ontologies represent semantics of Al-Quran as hierarchical concepts (Beirade et al., 2021; M. Yunus et al., 2017; Sherif & Ngonga Ngomo, 2015; Ta'a et al., 2014). Concepts or entities in Al-Quran are extracted whilst semantic relations between these Quranic concepts are defined based on hierarchy or taxonomy. However, this does not include semantics which are non-hierarchical such as meanings in events or stories that are told in a sentence. The semantics of events and stories can be represented through semantic relations such as predicate-argument or subject-verb-object relations (Jurczyk & Choi, 2015). Al-Quran contains a lot of stories with various didactic lessons (Salehi, 2016). Therefore, semantic relations between words in the Quranic verses should be represented as well.

Semantic graph representation has been shown to be a viable approach for representing knowledge of Al-Quran (Jurczyk & Choi, 2015; Noordin et al., 2016;