The Overview of Summary Interpretation Verses In The Holy Qur'an Based On Knowledge Graph Method

1st Khodijah Hulliyah[1], 2nd Rosmaria SW[1]
{khodijah.hulliyah@uinjkt.ac.id[1], rosmaria.widjajanti@uinjkt.ac.id[1]}

UIN Syarif Hidayatullah, Jakarta, Indonesia [1]

Abstract. To understand the texts in Qur'an, human are needed a good capability of language, linguistic and Tafseer al Qur'an. Why? Because understanding the verses the Qur'an could have much interpretation for each human. It depends on their knowledge. There the same verses of the Qur'an are still ambiguous meaning depending on how humans understand the verse fit their knowledge. Many methods are used to obtain effective and efficient results. This research, we analyze the summarization of text based on the knowledge graph concept. Knowledge graph method generates new knowledge gained from the summary text that has been grouped according to the type of relationship to each, using the concepts relate one word with another word in the parsing process, knowledge graph can decipher the ambiguity in interpreting a sentence. The output of this research is a summary of the verses interpretation in the graph formed.

Keywords: ontology, the summary of verses, concept, relationship, graph simplify, NLP, Text Summarization

1 Introduction

The Quran is the central religious text in Islam. Many aspects included in Quran, there are about lingustic, islamic law, devine direction and guidance and science and philosophy. All believers should learn text –translations and interpretations. To undestand of texts in Quran, human is needed a good capability of language, linguistic and Tafseer *al Qur'an.* Why? Because understanding the verses Quran could have much interpretation for each humans. It is depend on their knowledge. There the same verses of the Qur'an are still ambiguous meaning depending on how humans understand the verse fit their knowledge. In computer science, recently, text manipulation is became the favorite researchers. Scienties muslim have a big concentration to do the deep digging recited of Quran research. Knowledge Graph is one of the ontology methods in NLP that it works how to summary text of verses in Quran to understand and interprete the meaning text to get a new knowledge. Basically, knowledge graph theory uses two basic terms, namely concepts (tokens and types) and relationships (binary and multivariate relations) and the output of knowledge graph theory is a graph formed.

Text summarization is a field of NLP, artificial intelligence, and linguistic concerned with the interactions between computers and human languages. Summarization is a very useful method for many fields that develop natural language processing tasks.

The overview of Artificial Intelligence (AI) and Natural Language Processing (NLP) are: an overview of how NLP works is NLP has input and output, which includes the input that can be texts that come from various sources, while the output is in the form of knowledge base and data storage.

However, there is quite less research about the use of text summarization of Islamic studies together with the technology in one of the artificial intelligence branch, which is the Natural Language Processing, especially summary of text. Quran is the main source of knowledge and has been a major source reference for all types of problems.

To overcome this problem, we conduct research to gain understanding in reading the intentions contained in the verses of the Qur'an. Therefore, the objectives we will do are:

1. To analyze whether knowledge graph can be used to identify words that have a one-word relationship with other words.
2. To build a mind map concept in repetitive verses to get new knowledge from a collection of related words between this ayat and previous/after it.

Futhermore, this work, could answer these research question:

1. How to get words that have interrelated relationships in one verse with the other verses trough preprocessing stage?
2. Does the Knowledge Graph method can be used to get words that are interconnected in the form of graphs?
3. What is the contribution given by using knowledge Graph concept in islamic studies?

2 Related Works

The summarization text system is used to summarize sentences based on certain criteria. however, the results obtained, sometimes still often experience loss of important information in it. Therefore, based on Dimas's paper[1] that, a system is needed to get text abstraction, which represents all the information contained in the original text. Knowledge charts offer ways to build such systems.

Another research which has the same idea is how to get the abstraction summarization using the semantic graph reduction approach [2]. This study explains the process of making an abstractive summary for single document using semantic graph reduction technique. So, the step taken are to create a semantic graph according to the original document, then through several processes, a summary of the abstraction will be obtained, in the form of a graph that is simpler than the original graph.

Adopting a graph-based approach is used by Kavita, et.al. as well. This work has a novel graph-based summarization framework (Opinosis). They present a novel technique that produces brief abstraction summaries of highly redundant opinions.

In addition, there is the research of text mining for the processing text of the holy Qur'an. Al Hawarat paper [3] is to find an approach for analyzing Arabic text. It calculate the term of frequencies using both Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) methods.

3 Text Summarization

Information Technology is a part of the advancement of computer technology in the field of information at this time. The computer as a tool that has the structure of the machine follows the analogy of the human senses. Output in accordance with the tongue and the human hand. Processor and memory in accordance with the human brain and controls as well as a program corresponding to the human brain. Thus the process of informatics in the computer is an extension to the thought processes in the human brain. A computer is just a tool that helps in the processing of information[4].

The core of this technology is text processing. The text is to be read, understood and processed in order to be useful information into a knowledge. Have the means to deliver human knowledge becomes wise. This is the pinnacle of human endeavor to be a perfect man, where's the journey of the man in this world moves from phase to phase as revealed in the Qur'an starting from the data, information, knowledge, and wisdom. ICT as al-Qalam assists Humans to reach the highest level of this hirarchy the which is Wisdom[5].

A Summary can be defined as text that is generated from one or more texts, which contains most of the information in the original text (s), and that is not more than half of the original text [6]. summary sources can be one document or a combination of several documents. According to [7] the summarization process must not remove important information from the original determination. when the process is done manually, the results may be better, but if the file must be summarized on a large scale, it certainly requires a text summarization tool using a computer, which is automatically. this is called automatic text summarization. The output of the summary system can be in the form of extracts (ie when the choice of the "significant" sentence of the document is done) or abstract, when the summary can function as a substitute for the original document [8].

In computer science, recently, text manipulation becomes the favorite researchs. Scienties muslim have a big concentration to do the deep digging recited of qur'an researchs. One of them is how to summary text in Qur'an to understand the meaning of verses qur'an. Why it can be? According to verse Ali-Imran; 7, that sound is: That in interpreting the verses of the Qur'an, sometimes they translate according to their own will. Then you must have a belief that no one knows the interpretation except Allah. we believe that all is from God.

Behold, there are verses of the Qur'an are still ambiguous depending on how humans understand the verse fit their knowledge

4   Theory of Knowledge Graph

Knowledge graph (KG) is a new theory in the field of text summarization. KG is used to describe human language abstraction and extraction understood by computers. the approach taken on this theory is more to the semantic aspect approach than the syntactic aspect. therefore the concept of ontology in text summarization using KG theory is indispensable, or in NLP, included in the semantic network category. in principle, the composition possessed by KG is a concept (vertexs and types) and relationship. Knowledge graph method generates new knowledge gained from the summary text that has been grouped according to the type of relationship to each, using the concepts relate one word with another word in the parsing process, knowledge graph can decipher the ambiguity in interpreting a sentence.

Knowledge graph theory is a a new approach that can be used to express human language in graph formed. The fundamental difference between knowledge graph theory and other summarizing theories is that the theory of knowledge graph uses limited number of ontologies

or relations. Knowledge graph theory is able to describe or describe more basic semantic aspects, using a number of limited relations. This theory provides a new way of doing research to understand human language with the help of computers [9].

The development of information technology captures the phenomenon, that many things that will be encountered if the individual has a different perception and interpretation of the existing texts, such as legal text or texts in the Qur'an. So scientists in the field of computer soft side trying to do research to create new algorithms to perform summary of the texts are ambiguous. This method is known as the knowledge graph, where the technique is trying to do an analysis of the text to interpret the meaning of the sentence.

Since the artificial intelligent development was carried out, computer technology has been so sophisticated, that many human works can be done by computers, including the introduction of natural language. Generally, human natural language is the method most widely used to express human ideas and convey information. However, between formal languages that are understood by computers differ from natural languages. There is a gap between formal language (used by computers) and natural language. Communication between computers and humans is only possible when a lot of research is aimed at bridging this gap. An approach to bridging the gap is often called natural language processing (NLP), natural language understanding, or computational linguistics [9].

Knowledge graph is one of the references in NLP to represent a new way of describing natural language and modeling, and also makes a big step forward towards semantic understanding "know it and know why". Knowledge Graph means getting things not strings[10]:

The concept is important in shaping an understanding of specifically to the public or vice versa [2]. The concept can be divided into three types, namely token, type, and name Token is concept that are understood by one's way of view respectively, so this token is subjective. A concept relates to the meaning of the word; Type is a concept in the form of information general and objective because it is an agreement made beforehand. Examples of types such as fruits, animals and so on; and name is something that is individual.

4.1 Knowledge Graph Ontologies

An ontology is the naming and formal definition of types, properties, and reciprocal relationships of entities that are truly or fundamentally present for a particular domain of discourse. Thus the practical application of philosophical ontology, with taxonomy. An ontology squares the variables needed for several sets of calculations and establishes the relationship between them [11]. There is a philosophical review, ontology is the study of something real. Ontology is a theory that can explain the meaning of an object, the property of an object and the object relation that might occur in a knowledge domain. Ontology is based on two parts, the definition of concepts and the relations among them. Many ontologies aspect in the knowledge graph are token (concept) and 8 relationships of types.

Word graph ontology to date consists of tokens declared with nodes, 8 binary relationships, and 4 frame relationships. These eight binary relationships are [12]: Causality (CAU); Equality (EQU); Subset (SUB); Alikeness (ALI); Disparateness (DIS); Ordering (ORD); Attribution (PAR); Informational dependency (SKO).

5 Research Methodology

In this point several phases are discussed conducted in this research:

a. Literature Study Documents of the Qur'anic verses in Indonesian translation, This activity was carried out to collect the repeated verses needed in the study, we did a simulation on the recurring verse on surah Ash Shu'ara.

b. Determination of Nouns as Concepts, nouns are selected from each Indonesian text based on their characteristics. Then the nouns that have been selected are counted and grouped based on the similarity of the meaning of the word or the general word form. In this stage, we need the threshold.

c. Graph Making, nouns that have been determined as concepts will be labeled and used as a vertex to make a directed graph according to the intervertex relationship that occurs in each sentence based on the knowledge graph method.

d. Graph Analysis, at this stage an analysis will be carried out on each text by comparing all the graphs formed in each sentence and consider the relationships that occur between vertices which is there to determine the relationship between graphs.

e. Determining the Rules At this stage a rule will be used to summarize a translation of the text in the Qur'anic verse.

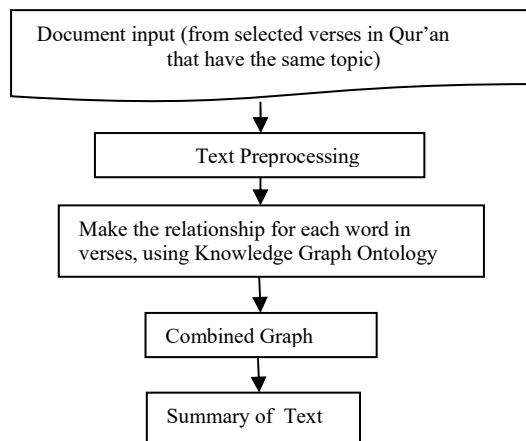Overall the overview of Research Design is could see in figure 1.

Document input (from selected verses in Qur'an that have the same topic)

↓

Text Preprocessing

↓

Make the relationship for each word in verses, using Knowledge Graph Ontology

↓

Combined Graph

↓

Summary of Text

Figure 1. The overview of research design

For the third step in the research design, we adopt knowledge graph method.

The following od the knowledge process

This paper will incorporate the concept of text summarization to perform the extraction text in verses. The process to obtain the summary of text that we use the KG Ontology approach. The following is the research design using Knowledge Graph method.

With KG, the texts are analyzed by:

1. Determine the concepts written in the text
2. Determine the relationships between the concepts mentioned above.

3. The concepts and relationships that are formed are translated literally in the form of graphs called text graphs
4. In KG, it is also necessary to interpret the text based on the background information that someone has (background knowledge) and also translated in graph form into text graph as well.
5. From both graphs (literal graph and background graph), combined into a new graph (combined graph) into a new text graph.
6. The new text graph is certainly very complicated to read and understand because it involves many concepts, it is necessary to simplify the graph.
7. With the formation of a simplified graph, the most important concepts in the text can be identified, resulting in a conclusion.

The previous step before used knowledge graph is we collect repeated verses accompanied by the preceding and following verses. Furthermore, we do pre-processing to get the words. The following stages of preprocessing are in fig.2
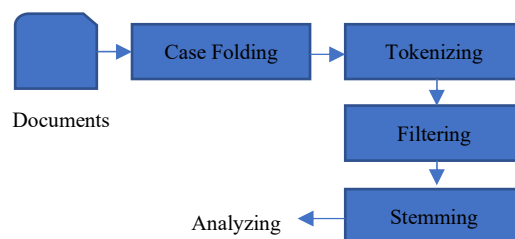

Figure 2. The Preprocessing Stages

6 Discussion

6.1. Literature Study Documents of the Qur'anic verses in Indonesian translation

We took on example In the quran, there are repetitions up to 8 times in Surah 26 Ash-Shu'ara (Poets): 9, 68, 104, 122, 140, 159, 175, 191, that is: وَإِنَّ رَبَّكَ لَهُوَ الْعَزِيزُ الرَّحِيمُ

Moreover, we analyzed based on the translation of the Qur'an (in Bahasa), to get a picture of the meaning of the repeated verse based on the context contained in the verses before and after it. The following are the ayah number 8, 9 and 10:

"dan sesungguhnya Tuhanmu benar-benar Dialah yang Maha Perkasa lagi Maha Penyayang";

and the previous verse is *"sesungguhnya pada yang demikian itu benar-benar terdapat suatu tanda kekuasaan Allah. Dan kebanyakan mereka tidak beriman"*

next, for the verse after that is: *"Dan ingatlah ketika Tuhanmu menyeru Musa dengan firmanNya, datangilah kaum yang zalim itu"*

6.2. Determination of Nouns as Concepts

The concept used in this work is in noun words. The noun used is taken from each sentence by considering the meaning of the formed sentence structure.

Table 1. The List of Noun Groupings and Their Number of Occurrences

Words	Total
Tuhanmu	2
Dialah	1
Maha Perkasa	1
Maha Penyayang	1
Tanda kekuasaan	1
Allah	1
Mereka	1
Tidak beriman	1
Musa	1
firmanNya	1
Kaum zalim	1

6.3. Graph Making

Before creating the graph, we first determine the vertices:

Table 2. Vertices

Vertex	Word	Total
v1	Tuhanmu	2
v2	Dialah	1
v3	Maha Perkasa	1
v4	Maha Penyayang	1
v5	Tanda kekuasaan	1
v6	Allah	1
v7	Mereka	1
v8	Tidak beriman	1
v9	Musa	1
v10	firmanNya	1
v11	Kaum zalim	1

Furthermore, after obtaining vertices resulting from grouping nouns, concepts and vertices will form graphs. Firstly, a word graph will be created from each noun in sentences based on noun phrases. Then the formed word graph will be matched with a predetermined vertex. In the formation of graphs will be considered inter-fetish relationships, as well as the conjunctions that exist in each verses (before or after).

The previous verse is:

"dan sesungguhnya Tuhanmu benar-benar Dialah yang Maha Perkasa lagi Maha Penyayang". From this sentence the phrase is obtained objects as follows: (a) *Tuhanmu;* (b) *Dialah;* (c) *Maha Perkasa*; (d) *Maha Penyayang*

The recurring verse is:

"sesungguhnya pada yang demikian itu benar-benar terdapat suatu tanda kekuasaan Allah. Dan kebanyakan mereka tidak beriman", Then, for the object of sentence phrases are: (a) *kekuasaan;* (b) *Allah;* (c) *Mereka;* (d) *Tidak Beriman*

And the last verse is:

"Dan ingatlah ketika Tuhanmu menyeru Musa dengan firmanNya, datangilah kaum yang zalim itu". The object of sentence phrases are: *(a) Tuhanmu; (b) Musa; (c) FirmanNya; (d) Kaum zalim*

After obtaining the concepts of each word of sentence object, we made the words graph form. See in figure 3

Figure 3. Words Graph

6.4 Graph Analysis

Then, these three verses, we are looking for important words that provide information about the meaning of the verses by going through the preprocessing process. For the main ayah, we got the result of preprocessing the related words, namely: *tuhan, dia, maha perkasa, maha penyayang;* for the previous ayah, namely: *tanda, kekuasaan, allah, mereka, iman;* and next the after ayat, namely: *ingat, tuhan , musa, kau, zalim*

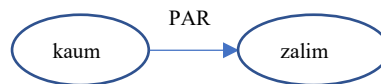If noun phrases are described in the graph, it is obtained:

Figure 4. The enhance for the phrase of "*kaum zalim*"

The result is a description of the words related to the recurring verse as follows in fig. 5
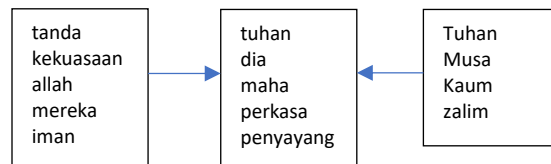

Fig. 5. The result of Preprocessing Step

The procedure for transforming text into KG to get a combined graph, you must go through preprocessing first. The follows is:

1. Text analysis, the text is read carefully to identify concepts and relations literally according to the text, to text concepts and relations
2. Adding knowledge background to background concept and relations.
3. Both texts (literal text and background text) are combined to form a combined graph.
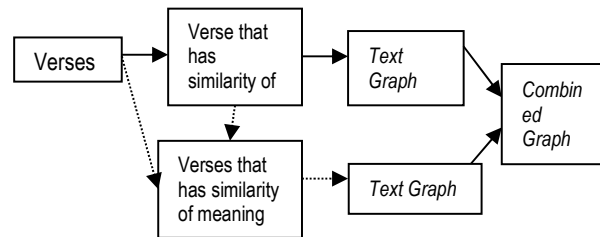4. The combined text is implemented into the software created.

Figure 6. Combined Graph

Simplify step One of the most fundamental things in making graphs is how to make problems that have been translated into graphs, become easier to understand or easier to read. Therefore the next step is to simplify the graph (Simplified Graph) Simplification of graphs is not only useful in limiting the number of concepts and relationships, but also helps solve problems related to synonyms and specialization of these concepts [9]

Simplification process of graphs, which includes:
1. Formation of general concepts (general concept)
2. Identification of constructs (identifying constructs)
3. Establishing an independent subgraph (identifying   identity subgraphs)
4. Removal of causal chains (reducing causal chains)

To identification of contruct, KG have eight binary relationship. The 8 binary relationship describe: equ, sub, ord, cau, par, ali, dis, and sko concepts.

Finally, we find the result of the simulation of summary verses interpretation, as shown in fig.6
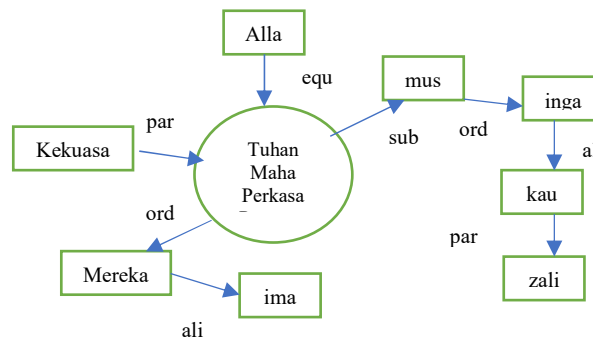

Figure 7. The Result of the simulation of summary verses interpretation

7   Conclusion

The conclusion of this work are:

1. This research is still at the level of analysis whether the preprocessing stage and knowledge graph method can be used to build a database with mind map concepts that can be used as a reference in facilitating understanding of the relationship of a verse with the preceding and following verse.

2. This research is still not perfect, because the results obtained have not been compared with other techniques, so it cannot be said as optimal results.

3. Allah commands to read and write, is the foundation of man's exploration and in-depth research on the phenomena of the universe for the sake of human survival. With the provision of reasonable and mind, then human knowledge continues to grow, accompanied by the advances in technology that can help answer the question of what happened.

References

[1]   D. Febriatmoko, D. Matematika, F. Matematika, D. A. N. Ilmu, and P. Alam, "ABSTRAKSI TEKS BERBAHASA INDONESIA," 2011.

[2]   I. F. Moawad and M. Aref, "Semantic graph reduction approach for abstractive Text Summarization," *Proc. - ICCES 2012 2012 Int. Conf. Comput. Eng. Syst.*, pp. 132–138, 2012.

[3]   M. Alhawarat, M. Hegazi, and A. Hilal, "Processing the Text of the Holy Quran: a Text Mining Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 2–7, 2015.

[4]   D. Kemenag, *Islam Untuk Disiplin Ilmu Teknologi*. 2004.

[5]   F. M, *ICT and Islam*. IIUM Printing, 2013.

[6]   E. H. Hovy, "Language Production, Cognition, and the Lexicon," *Springer Int. Publ. Switz.*, pp. 13–24, 2015.

[7]   I. Mani, "Summarization evaluation: An overview," *Nat. Lang. Process.*, p. 286, 2001.

[8]   D. Das and A. Martins, "A Survey on Automatic Text Summarization Single-Document Summarization," *Not Publ.*, pp. 1–31, 2007.

[9]   L. Zhang, *Knowledge graph theory and structural parsing*. 2002.

[10]   F. Li, "Things , not Strings : Query Analysis with Knowledge Base," 2012.

[11]   A. Blumauer, "Semantic Web & Linked Data in Enterprises An overview Welcome !," 2015.

[12]   Sri Nurdiati and Cornelis Hoede, "25 YEARS DEVELOPMENT OF KNOWLEDGE GRAPH THEORY : The Knowledge Integr ation and Str uctur ing System (KISS)," *Knowl. Creat. Diffus. Util.*, pp. 1–10, 2008.