



Development Grouping of Synonym Set Thesaurus Vocabulary of The Qur'an in English using Hierarchical Clustering Algorithm

Salma Fauziah^{1*}, Moch Arif Bijaksana²

^{1,2} Faculty of Informatics, Telkom University, Bandung, Indonesia

^{1,2} Jl Telekomunikasi No.01, Dayeuhkolot, Bandung, Indonesia

*Corresponding email: salmaf@student.telkomuniversity.ac.id

Received 10 May 2020, Revised 27 May 2020, Accepted 01 June 2020

Abstract — Research in the field of text mining to process entries or words from the Qur'an is very beneficial for Muslims. However, the study in this particular area is quite limited. Therefore, this work aims to establish a set of synonyms for the thesaurus in the words of the Qur'an. The employed dataset is corpus Qur'an and English translation. This research improves the previous work done by Laras Gupitasari, namely "The Development of Al-Qur'an Vocabulary Set Synonyms with WordNet Approach." We use nouns from the translation of English words in the Quran as an input, while the system produces several groups with the same level of closeness of meaning displayed as the output. This study uses the word hierarchical grouping method and calculates distances using common paths to produce output. The groups have resulted in the closeness of meaning from word entries. The first group of the output means it has the closest meaning. The evaluation to measure the accuracy of predictions issued by the system is called an F-Measure by 76%.

Keywords – thesaurus, synonym set, text mining, hierarchical clustering

Copyright © 2020 JURNAL INFOTEL

All rights reserved.

I. INTRODUCTION

The Qur'an is an eternal miracle in Islam with increasingly advanced knowledge and technology today [1]. The Qur'an consists of 114 Surahs, 30 Juz, and 6236 Verses. Each verse of the Qur'an has a different meaning [2]. Every verse in the Qur'an has the same meaning, among other verses, called the synonym set. A synonym set is a collection of one or words that have the same meaning or synonym. Every word with the same meaning can replace part of the word used without changing the meaning or meaning of the sentence [3]. Indonesian also has a database set up as a website to find synonyms with Indonesian thesaurus references [4]. Valid synonyms in WordNet only work for two words that have a synonym relationship. If the two words do not have a synonym relationship, the two words are invalid [5].

The synonym search system of similar words is WordNet, but the WordNet system found only uses an English search [6]. WordNet contains not only

synonymous words but also contains opposite meanings [7]. Not only is WordNet able to contain words that have the same meaning, but Indonesian also has a book of synonyms of two or more words that have the same meaning. Thesaurus contains language vocabulary that corresponds to a close relationship. Thesaurus can make it easier for readers to find words [8]. Thesaurus can help indexers look for descriptors for keywords from documents [9].

Clustering algorithms like hierarchical clustering can be one algorithm that can build synonym sets with data sets derived from the verses of the Qur'an. The hierarchical clustering algorithm is chosen based on its function to form clusters of each data point and build clusters by distinguishing each object selected in the top-down or bottom-up model. The hierarchical clustering algorithm's advantage is that it can provide a grouping of data using distance measurements. This is very suitable to be applied to measure the distance between words in a data set. It also performs a

hierarchical grouping that can build a set of synonyms that previously unknown about how many synonyms will be found [10]. It was once compared with the K-Means clustering algorithm, but it was not suitable for this dataset in finding a set of synonyms. Because in the beginning, it must specify K in K-Means clustering and find out how the system will generate many synonym sets.

This research will apply hierarchical clustering to build synonym grouping sets using English and Arabic. The entire Qur'anic dataset is generated from openburhan.net. The results from the synonym set will be validated using the "gold standard". Gold standards are a type of evaluation that is needed in aligning the results of computer programs. Gold standards are useful for expert opinions that can be accepted as accurate references [11]. This study aims to produce a grouping of synonyms using the hierarchical clustering algorithm with English word input and Arabic output and to improve the accuracy of clustering results. A collection of synonyms produced by the system is a word that has the same meaning as the word entered. The system will then group to filter results into Arabic only and make it into several levels formed by a cluster. The clusters shown mean that the first group that appears means the group's words have a close meaning and the next group that appears means that the similarity distance is not too close. The purpose of this study is based on existing problems, namely the number of studies to build synonym sets, but no one has produced synonym sets using the synonym set

grouping method. The benefit of grouping synonym sets is that they can get synonym results based on the proximity of their meaning from the words entered. This research has a limitation that is only processing noun vocabulary.

This research is a follow-up study of an article titled "Development of Synonyms of the Qur'an Vocabulary with WordNet Approach" by Laras Gupitasari [12]. The previous article's difference is that this article develops a new process; namely, the synonym grouping set. It aims to group words from a group that produces the same meaning synonyms, and the results will be grouped in Arabic only. They will display Arabic groupings according to the proximity of meaning with the words entered. This article uses English input and Arabic output. In contrast with [12], that only uses Arabic input and English & Arabic output. This article also improves the clustering results, which means that the previous article's clustering results are not to generate a lot of related words.

II. RESEARCH METHOD

A. General Description of System

The system built aims to display several words that have the same meaning as the words entered in the noun class. The synonym input results set is processed through word embedding with hierarchical clustering. Fig.1 shows a general description of the system that will be built in this research.

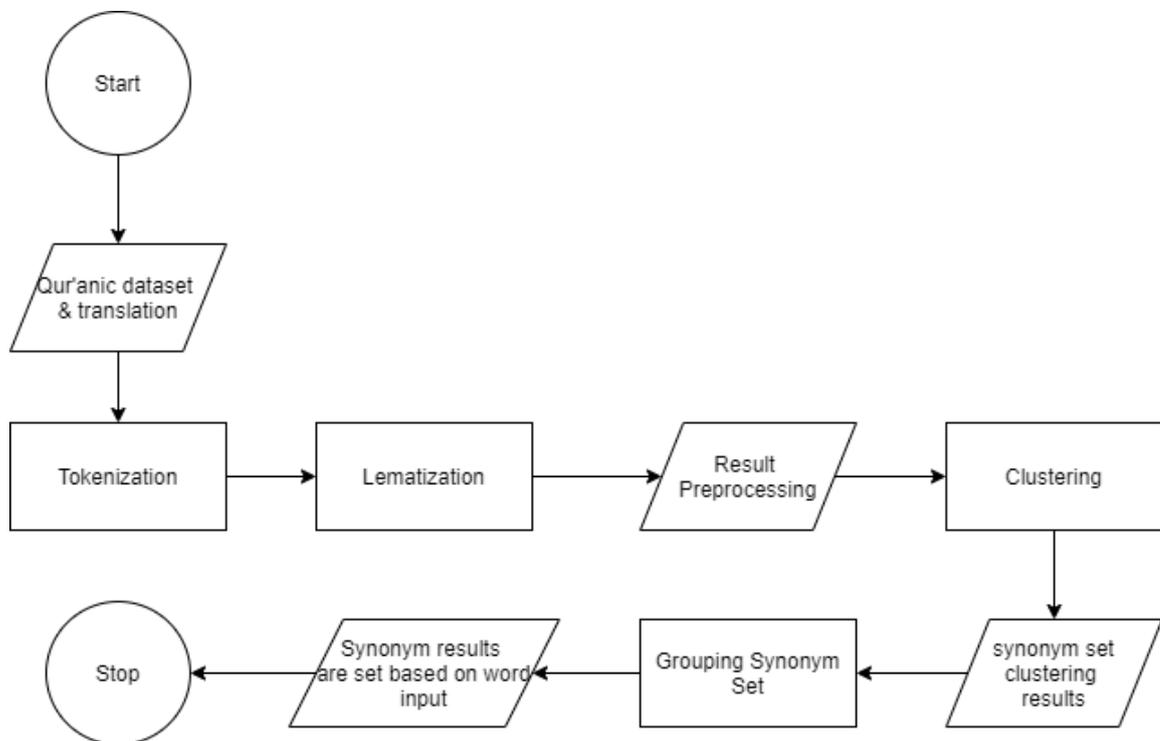


Fig.1. General Description System

B. Dataset Qur'an

The dataset processed was 77,795 data lines with 5,615 words. The number of words is in the Qur'an. The dataset used in this study is the results of searching data from the online site openburhan.net as well as some additional uses of the Qur'an corpus. The dataset in Latin written form/translation of the word Al-Qur'an in English will be used as input in the system, while the output of the system is in the form of Arabic words that have the same meaning as the words entered.

The system generates additional data, as many as 3,904 words have a synonym set of more than one element. There are 38 words with a synonym set consisting of only one element. With the process of grouping synonym sets that have been done at the end of the study, the system produces 454 words that have the same meaning from 50 test data with 364 valid words based on testing by the gold standard with the results of grouping synonym sets of 109 valid synset groups. Gold Standard also provides a number of words that he thinks are accurate and can have the same meaning as many as 51 words.

C. Preprocessing

Preprocessing is used to cleaning up raw and normalized data into a ready-use data [13]. This study uses processing to clean data from openburhan.net, so only the data needed in this study is ready for use. This Preprocessing phase will include two processes, namely lemmatization and tokenization. Lemmatization is the process of eliminating excessive translation words to produce basic words. The process of lipatization is followed by merging to combine each translation's basic words, which has the same meaning [14]. The results of the explosive process are used as entries in the system. Fig. 2 contains the lemmatization results and the merged process of lemmatization.

Table 1. Lemmatization and Merge Result

Lemmatization Result	Merged Result
['Ramadhan']	[رمضان, Ramadhan]
['Owner']	[أهل, Owner]
['Judgment', 'Conclusion', 'Opinion']	[حكم, Judgment], [نيط, Conclusion], [بدو, Opinion]

D. Tokenization

Tokenization means separating each word from a sentence. This study only found one word in English, which has many meanings. A tokenization process is needed to get the one word required to be processed [14]. All words that have been obtained will be processed by the system to produce words that only belong to the class of nouns. Table 1 contains some examples of the tokenization process.

Table 2. Example Tokenization

Before Tokenization	After Tokenization
['The Book', 'A Book', 'Of The Book', 'In The Book']	['Book']
['On The Day', 'And The Day', 'The Days', 'That Day', 'This Day']	['Day']
['Our Evil Deeds', 'Their Evil Deeds', 'And Evil It Is', 'Their Evil', 'An Evil Thought']	['Evil']

E. Clustering

After the data has been completed, the data will be processed using one of the clustering methods. Clustering is a method for organizing groups of objects that have the same characteristics. The purpose of grouping is to collect data into clusters so that some similarities between intra-cluster and inter-cluster are collected [1], [15]. There are many clustering types, but this study will use hierarchical clustering to combine words that have the same meaning of the words entered. Fig.2 below illustrates how hierarchical clustering works.

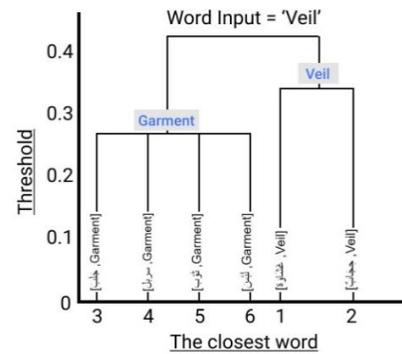


Fig.2. Illustration Hierarchical Clustering

When the system gets the input word 'veil', it calculates the input word's distance with all the words in the dataset. When the system gets a word with a threshold <= 0.5, it immediately makes a cluster with the same meaning based on the resulting threshold. This clustering technique uses the distance similarity calculation on WordNet to calculate the shortest distance from one word to another. It is a similar path to calculate the path between words that take into account conjunctions such as funds.

$$\text{SynonymPath}(C1,C2)= \frac{1}{\text{path}(c1,c2)} \quad (1)$$

In (1), c1 is the first word compared to the second word c2, with the maximum generated value is 1. After getting the results, the proximity value of the distance between words will be compared with the threshold of 0.5. The threshold is the limit or minimum value for the distance between words, with 0.5 is a fair limit number. By 0.5 of the threshold values, the results

produced to measure the distance between words are relatively stable and suitable for finding the closest meaning between two words. Threshold is used to get good synonym results. We have used the different minimum threshold value, but the synonym results show are inaccurate. There are some words that, in fact, a synonym, but does not match because the distance is too large—the threshold working method in this grouping. If the Proximity value is the same as the threshold or greater, it means the distance between words has a large proximity value. Words with a considerable closeness will be combined into a one-unit cluster so that one cluster will contain more than one word that has the same meaning. Table 3 contains the algorithm used in this system.

Table 3. Algorithm Clustering

Algorithm 1: Clustering synonym set	
1:	Input: Dataset Arabic Word & Translation
2:	For i to Arabic Words[i]
3:	For j to Arabic Words [j]
4:	Calculation path similarity between word I and j
5:	If distance >= Treshold then
6:	Insert two arabic word into one cluster
7:	Else
8:	Insert word in different cluster
9:	End
10:	Update each set every word update
11:	Output: Synonym Set

F. Grouping Synonym Set

The synonym grouping set's workings are that the system will take a group containing the English and Arabic non-language from the input word. For example, the input word is 'name,' then the system will issue words that have the same meaning of the word 'name' based on clustering results, and the system will remove synonym results from the name 'brand and 'people'. Next, the system will sort the results starting from the word input. If the system has a set of synonym sets sorted according to the same meaning, the system will divide into several groups to distinguish between words.

Following is an example table of synonym results set before grouping and after grouping. The table below compares the system's results by inputting the word 'Veil' and the results by the system when the synonym output set of hierarchical clustering is formed by grouping. Before the grouping results, the system displays all words that have the same meaning as the word 'Veil' but when the system processes the synonym result of the word to veil to be grouped according to the proximity of meaning. The system produces two groups, which, according to the system,

both have the same meaning, but one of the groups is closer in meaning.

Table 4. Comparison Table of Grouping Synonym Set

Before Grouping	[[Veil, عَشَاوَة],[Veil, حَجَلْب],[Garment, حَجَلْب],[Garment, سَرِبَل],[Garment, ثُوب],[Garment, لَبِيْس]]
After Grouping	[[لَبِيْس, ثُوب, سَرِبَل, حَجَلْب, عَشَاوَة]]

G. Accuracy Calculation

The calculation accuracy will be done by comparing the system that generates synonym sets and calculations using the F-Measure method. F-Measure also involves two factors, namely recall and precision, which will include the gold standard or human opinion used as a reference for a successful system.

$$\text{Recall} = \frac{N_{all}}{N_{System}} \quad (2)$$

$$\text{Precision} = \frac{N_{all}}{N_{gold}} \quad (3)$$

$$F\text{Measure} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In (2), N_{all} is the number of words in the synonym set that are in the gold standard, while N_{System} is the number of words in the grouping synonym set produced by the system. In (3), N_{gold} is the number of words in the synonym set made by the gold standard.

III. RESULT

The synonym development research set of thesaurus using hierarchical clustering has issued a test result by a 454 Arabic lemma system from the input of 50 English words. The test results of this system will be validated by linguists. Then, the overall system results and gold standard results will be calculated using F-Measure to produce total accuracy. The recall is counting the number of words in the selected synonym set; meanwhile, Precision counts the number of words in the correct and selected set of synonyms. Table 5 shows us the test results by the system.

Table 5. Test Result by System

Word Entry	Output Grouping Synonym Set
Veil	[[لَبِيْس, ثُوب, سَرِبَل, حَجَلْب, عَشَاوَة]]
Name	[[وَسْم, قَبِيْس, سُمُو, آلِه, حَسَن]]
Believe	[[صَدَق, قَبِل, آمَن, ظَن]]
Punishment	[[صَعَد, عَذَاب, رَجَز, نَكَال, عُقْب, يَلَس, رَجَس, أَلَم, مَثَل], [نُكِر, ذَا, لَزَم, جَزَاء, عِقَاب, أَيْم]]
Fire	[[وَقِع, شُهَب, نُور, جَدِيم, حَرَق, سَعَر, سَمَم]]

Word Entry	Output Grouping Synonym Set
Darkness	[[سَوَادٌ، حَمَاً، دَهَمٌ، أَحْوَى، [،ظُلْمَةٌ، لَيْلٌ، غَسَقٌ، قَتْرٌ]]
Witness	[[نَظَرَ، رَاعَى، رَقَبَ، [،قَوْلٌ، [،خَوْضٌ، [،بَصُرَ، شَاهَدَ]]، [وَكَّلَ]]
Stone	[[صَخْرٌ، صَفْوَانٌ، [،حَجْرٌ، نُصْبٌ، حَصْبٌ]]
Opinion	[[نَبْطٌ، [،يَوْمٌ، دِينٌ، قَوْمٌ، حَكْمٌ، آلَهُ، فَتْحٌ، فَصْلٌ، [،بَدْوٌ]]، [رَأَى]]
Book	[[سَفَرٌ، كِتَابٌ، أُمَمٌ، زَبْرٌ، أَهْلٌ]]
Story	[[حَدِيثٌ، [،سَطْرٌ، لِهْوٌ، حَدِيثٌ، [،حَدِيثٌ، نَبَأٌ، قِصَصٌ، [،سَطْرٌ]]، [قِصَصٌ، حَدِيثٌ]]
Sound	[[سَلْمٌ، خَوْرٌ، صَدَقٌ، سَوِيٌّ، رَكْزٌ، هَمْسٌ، حَسَسٌ، [،صَوْتٌ]]، [خَفْتٌ، صَوْتٌ، صَرْرٌ]]
Rain	[[طَلَلٌ، [،عَصْرٌ، وَبَلٌ، مَطَرٌ، سَمُوٌّ، غَيْثٌ، وَدَقٌ، [،مَزْنٌ]]، [سَيْلٌ، [،صَوْبٌ]]
Animal	[[أَنْمٌ، [،بِهِمٌ، وَحْشٌ، [،مَوْتٌ، هَدْيٌ، سَبْعٌ، جَرَحٌ، كَلْبٌ]]، [دَبَبٌ، ذَا، [،نَكَرٌ، سَرَقٌ، عَيْدٌ]]، [أَنْثٌ، ثَنِيٌّ، [،أَبْرَأٌ، دَبَبٌ]]
Journey	[[سَيْرٌ، ضَرْبٌ، طَعْنٌ، سَرِيٌّ، [،حَجَجٌ، كَبْرٌ، [،رَحَلٌ]]، [سَفَرٌ، سَيْرٌ]]، [ضَرْبٌ]]
Measure	[[رَبَبٌ، وَطَأٌ، [،وَقِيٌّ، حَنْزٌ، [،كُونٌ، حَسْبٌ، كَيْلٌ، قَدْرٌ]]
Friend	[[وَلِيٌّ، خَلَلٌ، عَشْرٌ، صَدَقٌ، حَمَمٌ]]، [كَعْبٌ، صَحْبٌ، قَرْنٌ، رَفَقٌ، قَصْرٌ، طَرْفٌ، حَوْرٌ، عَيْنٌ]]
Owner	[[عَرَشٌ، أَوْلٌ، مَلِكٌ، أَهْلٌ، ذَا، نَقْمٌ، طَوْلٌ]]
Protector	[[قُوْتٌ، خَزَنٌ، صَحْبٌ، [،حَفْظٌ، وَوَلِيٌّ، قَوْمٌ، عَصْمٌ، [،وَقِيٌّ]]، [قَوْمٌ، [،وَكَلٌ، عَصْمٌ، وَوَقِيٌّ، [،بَلْغٌ، [،حَفْظٌ، حَرَسٌ]]، [رَصَدٌ]]، [وَلِيٌّ، هَمَنٌ، حَفْظٌ، وَكَلٌ]]
Evil	[[سَوَاءٌ، بَأْسٌ، خَبِيثٌ، طَيْرٌ، رَجَزٌ، رَجَسٌ، شَرٌّ، نَكَرٌ]]، [غَوِيٌّ، [،غِيٌّ، عَدُوٌّ، شَطَطٌ]]، [ذَنْبٌ، جَرْمٌ، [،فَجْرٌ، فَسَقٌ]]
News	[[خَبْرٌ، نَبَأٌ، [،بَشْرٌ، رَيْبٌ، آلَهُ، [،حَدِيثٌ، بَشْرٌ، نَبَأٌ]]، [خَبْرٌ]]
Life	[[بَشْرٌ، أُنْسٌ، [،أَوْلٌ، حَيِيٌّ، عَمْرٌ، نَفْسٌ، دَنُوٌّ، بَدْوٌ]]، [عَيْشٌ، هَزْزٌ]]، [أَبْرَأٌ]]
Water	[[غَسَلٌ، مَوْهٌ، سَقِيٌّ، حَمَمٌ، وَرْدٌ، عَيْنٌ]]
City	[[أَفْكٌ، مَصْرٌ، بَلَدٌ، قَرِيٌّ، مَدَنٌ]]
Reward	[[آلَهُ، حَسَنٌ، أَجْرٌ، ثَوْبٌ، جَزِيٌّ، زَيْدٌ]]
Child	[[بَيْتَمٌ، غَلْمٌ، [،طِفْلٌ، بَنِيٌّ، وَوَلَدٌ، نَزْرٌ، كَلَلٌ، صِلْحٌ]]، [صَبِيٌّ]]
Abode	[[دَوْرٌ، بَيْتٌ، سَكَنٌ، [،أَوْبٌ، أُمَمٌ، ثَوِيٌّ، أَوِيٌّ، قَرْرٌ]]
Soul	[[قَدَسٌ، رَوْحٌ، [،عَلِيٌّ، نَفْسٌ، وَوَفِيٌّ، عَقْبٌ، رَوْحٌ]]
Sign	[[يَقِيْنٌ، أَيْبِيٌّ، بَيْنٌ، آلَهُ، بَصْرٌ]]
Desire	[[كَوْنٌ، هَوِيٌّ، رَوْدٌ، مَنِيٌّ، هَمٌّ، شَهْوَةٌ، غَيْرٌ، أَوْلٌ، أَرْبٌ]]، [رَجْوٌ، طَمَعٌ]]، [فَادٌ، [،رَوْدٌ، [،رَوْدٌ، مَنِيٌّ، آلَهُ]]
Strength	[[بَدِيٌّ، قَوِيٌّ، طَوْقٌ، جَبْرٌ، رَوْحٌ، مَحَلٌ، أَزْرٌ، شَدْدٌ، أَوْلٌ، ذَا]]، [أَيْدٌ،

Word Entry	Output Grouping Synonym Set
	[[قَوِيٌّ، [،صَبِيْرٌ]]
Love	[[بَتَلٌ، عَكْفٌ، [،وَدٌ، حَبِيْبٌ]]
Permission	[[رِضْوَانٌ، [،أُنْسٌ، أَذْنٌ، رُوْدٌ، آلَهُ]]
Harm	[[غَسَلٌ، قَرَحٌ، جَرَحٌ، [،ضُرْرٌ، سَوَاءٌ، فِتْنٌ، أَذِيٌّ، ضَيْرٌ]]، [عَرْرٌ، [،قَرَحٌ، [،أَذِيٌّ، ضُرْرٌ]]]
Possessor	[[ذُوٌّ، ذَا، أَوْلٌ، قَوِيٌّ، أَيْدٌ، رَفْعٌ، دَرَجٌ، بَأْسٌ، مَرْرٌ]]، [فَضْلٌ، [،قَرْدٌ، [،مَرَأٌ، نَفْسٌ، عَمْرٌ، فُسُقٌ، سَكَنٌ]]]
Helper	[[تَبَعٌ، [،عَمَلٌ، [،ظَهْرٌ، وَزْرٌ، [،إِرْدَأٌ، نَصْرٌ، صِرْحٌ]]، [عَضُدٌ، ظَهْرٌ]]
Way	[[ضَلَلٌ، طَرِقٌ، سَوِيٌّ، نَهْجٌ، دَأْبٌ، قَوْمٌ]]، [فِيْرَشْدٌ، قَرَبٌ، صَرَفٌ، آلَهُ، سَبَبٌ، رَعٌ، خَرَجٌ، عَرَجٌ، نَجْدٌ، سَيْلٌ، نَسْكَ، وَوَسْطٌ، سَنَنٌ، [،صِرْطٌ]]، [حَكْمٌ، [،شَكْلٌ، عَرَفٌ، مَثَلٌ]]
Fruit	[[قَطْفٌ، ثَمْرٌ، أَكَلٌ، فَكْهٌ، طَلْعٌ، جَنِيٌّ، صَرْمٌ]]، [زَيْتٌ، [،نَقْرٌ، فَتْلٌ، حَبِيْبٌ، خَرْدَلٌ، زَرْعٌ]]
Sign	[[يَقِيْنٌ، أَيْبِيٌّ، بَيْنٌ، آلَهُ، بَصْرٌ]]
Party	[[نَفْرٌ، حَزْبٌ، فِءَاءِيٌّ، طَوْفٌ، فَرَقٌ]]
Force	[[عَصْبٌ، جَنْدٌ، كَرَهٌ، فَأْيٌ، جَمْعٌ، قَوِيٌّ]]
Neck	[[جَبَدٌ، رَقَبٌ، طَوْقٌ، عَطْفٌ، عَنَقٌ]]
Hear	[[سَمِعَ، [،شَعَرَ، أُنْسٌ، [،صَمَمٌ، سَمِعَ، مَوْتٌ]]]
Kindness	[[مَنْزَنٌ، [،مَعْنٌ، رَأْفٌ، حَسَنٌ، عَرَفٌ]]
Escape	[[فَرَرٌ، فَوْزٌ، حَيْصٌ، عَجَزٌ، عَزَبٌ، صَرَفٌ، وَوَأَلٌ، فَوْتٌ]]، [نَوْصٌ]]
Determination	[[أَفْكٌ، [،حَرْدٌ، أَمْرٌ، أَوْلٌ، عَزْمٌ، قَدْرٌ]]
Sacrifice	[[نَحْرٌ، نَسْكَ، قَرَبٌ، نَصْبٌ، حَلَلٌ، ذَبِيْحٌ]]
Land	[[بَرْرٌ، حَرْتٌ، بَلَدٌ، أَرْضٌ]]
Blaze	[[شَهَبٌ، مَرَجٌ، شَوْطٌ، لَهَبٌ، [،لَهَبٌ، حَجْمٌ، سَعْرٌ]]، [حَرَقٌ، لَطْيٌ]]، [شَهَبٌ، [،سَعْرٌ]]
Wealth	[[تَرَفٌ، [،خَيْرٌ، مَوْلٌ، قَنْطَرٌ، وَرَثٌ، طَوْلٌ، آلَهُ]]
Commit	[[جَرَحٌ، أَتَى، غَلَوٌ، قَرَفٌ، سَرَفٌ، زَنَى، عَثُوٌّ]]

IV. DISCUSSION

The dataset used in this study was 77,795 entries or Arabic words, processed by a system that produced nouns 5,615 entries. The system processes 5,615 entries to find the same meaning from the test data of 50 translated words. It makes 454 words that have the same meaning and several words, including Islamic words (words that only exist in the Qur'an) and subject words (words that mark someone's identity).

Each word entry from the system results has many outputs that are limited by arrays or delimiters that display groups of words with the same closeness meaning. This means that the result of the first system, which is the closest meaning in terms of the word 'veil' is the first group, namely *جَبَابٌ*, *غَشَاوَةٌ*, the next result means it has the same meaning as well but the

closeness of the meaning is far more than the results in the previous group.

Keep in mind that this article is a follow-up study of a journal that has been published by Laras Gupitasari [12]. The dataset and clustering method used in this study is the same as [12]. The two articles have a difference in the addition of the process of grouping synonym sets. It aims to issue several clusters based on the same meaning with the closest distance from the inputted word and making grouping from Arabic results. So, from the grouping, it can be seen that each word that has the same meaning has a level of closeness. In contrast to the previous article, they only outputs the full result from the synonym set. Here is a comparison picture of the process produced by [12] and this article when the system entered the word 'veil':

Result Article by Laras	Result article by Salma
[[Veil, غشاوة], [Veil, حجاب], [Garment, جلب], [Garment, سربل], [Garment, ثوب]]	[[لبس ثوب, سربل, جلب], [حجاب, غشاوة]]
[[Garment, ثوب], [Garment, لبس]]	

Fig.3. Process Comparison Between Previous Articles

The picture above explains in the red line that the two arrays are a synonym set of words entered, namely 'veil.' Synonyms of 'veil' involve six different Arabic elements. The set veil synonym results will be grouped or filtered by the system and produce only one Arabic cluster, but the English meaning is the same. Another difference is in the test data used. Table 6 below is a comparison final accuracy result.

Table 6. Comparison Table of The Accuracy Result

Research	Data Test	Precision	Recall	F-Measure
This Research	50	71.4 %	80.1 %	76%
Previous research version	7	86%	81%	83%

This study uses the gold standard test to validate the system's results and add a few words with the same meaning but from linguists. This study uses several reference sources, namely the Almaany Arabic Dictionary and the Munjid Dictionary for gold standard testing. This test was also carried out in a previous version of the research, already seen in comparison in Table 6. After being tested by the gold standard testing, the total results of a valid system were 364 words, while the test results from the gold standard alone produced 51 Arabic words. Both results were processed to obtain the accuracy of this study. Accuracy calculations using recall, precision, and F-Measure are in the appendix at the end of this paper.

Next is the presentation of pure synonym results set with the system, valid system results based on the gold standard and additional gold standards as well as tables

showing which words are invalid based on the gold standard.

Table 7. Comparison System Result and Gold Standard

Word Input	System Result	Sytem Valid + Gold Standard
Veil	لبس, ثوب, سربل, جلب, حجاب, غشاوة	لبس, ثوب, سربل, جلب, حجاب, غشاوة
Name	وسم, سمو, آله, حسن	وسم, سمو, آله, حسن, اسم, تسمية
Believe	صدق, قيل, آمن, ظن	صدق, آمن, ظن, اعتقد
Punishment	صعد, عذاب, رجز, نكال, عقب, باس, رخص, ألم, مثل, كثر, ذل, لزم, جزاء, عقاب, آثم	فصاص, عذاب, رجز, نكال, عقب, باس, رخص, جزاء, عقاب
Fire	وقع, شهب, نور, جسيم, حرق, سحر, سم	ثار, شهب, نور, جسيم, حرق, سحر, سم
Darkness	سواد, حملا, دهم, أخوى, ظلمة, ليل, غسق, قتر, عتمه	سواد, أخوى, ظلمة, ليل, غسق, قتر, عتمه
Witness	نظر, راعى, رقب, قول, خوض, بصر, شاهد, وكل	نظر, رقب, بصر, شاهد, وكل, بيته, يرها, دليل
Stone	صخر, صقوان, حجر, نصب, خصب	صخر, صقوان, حجر, نصب, خصب, حجري
Opinion	نيط, يوم, دين, قوم, حكم, آله, فتح, فصل, بدو, رأي	دين, قوم, حكم, فتح, فصل, رأي, فكر
Book	سفر, كتب, أمم, زير, أهل	سفر, كتاب, زير, دفتر
Story	فصص, حدث, حدث, سطر, لهو, حدث, حدث, نبا, قصص, سطر	حكاية, رواية, حدث, حدث, لهو, سطر, حدث, فصص, حدث, سطر, نبا, قصص
Sound	خفت, صوت, صرر, صدق, سلم, خور, صدق, سوي, ركز, همس, حسس, صوت	نعم, خفت, صوت, خوار, صدق, سوي, ركز, همس, حسيب, صوت
Rain	سيل, صوب, طلل, عصر, وبل, مطر, سمو, غيث, ودف, مزن	سيل, طلل, عصر, وابل, مطر, سماء, غيث, ودف, مزن, غيث
Animal	أنث, ثني, برأ, دبب, أنم, بهم, وحش, موت, هدي, سبع, جرح, كلب, دبب, ذل, ذكر, سرق, عيد	أنثى, ثني, برأ, دبب, للأتام, بهيمة, وحش, هني, سبع, جرح, كلب, ذل, ذكر, حيوان
Journey	ضرب, سير, ضرب, طعن, سري, حجج, كبر, رحل, سفر, سير	يضربون, سيز, يضربون, طعن, سيز, بالحج, رحل, سفر, سيز, رحل, رحلة
Measure	ربب, وطأ, وقي, حتر, كون, كيل, قدر	وطأ, جنز, حساب, كيل, قدر, حجم, قياس
Friend	كعب, صحب, قرن, رفيق, قصر, طرف, حور, عين, ولي, خلل, عشر, صدق, حمم	كواعب, صاحب, قرن, رفيق, حور, أولياء, خليل, عشير, صادق, حميم
Owner	عرش, أول, ملك, أهل, ذل, نعم, طول	عرش, أول, ملك, أهل, ذل, نعم, ذو
Protector	ولي, همن, حفظ, وكل, قوت, خزن	ولي, همن, حفظ, وكل, قوت, خزنة

Word Input	System Result	Sytem Valid + Gold Standard	Word Input	System Result	Sytem Valid + Gold Standard
	صاحب, حفظ, ولي, قوم, عصم, وفي, قوم, وكل, عصم, وفي, بلغ, حفظ, حرس, رصد	صاحب, حفظ, ولي, قوم, عصم, اتقى, قوم, وكل, عصم, وفي, بلغ, حفظ, حرس, رصد, مناصر	Helper	تابع, عمل, ظهر, وزير, ردا, نصر, صرخ, عضد, ظهر	تبعه, عمل, وزير, رده, نصر, مبرح, عضد, ظهور, معاون
Evil	طغي, عدو, شطط, سوا, باس, خبث, طير, رجز, رجس, شرر, نكر, غوي, ذنب, جرم, فسق	سوءة, باس, خبث, طائر, رجز, رجس, اشرار, نكر, غوي, هجر, تعنوا, ذنب, جرم, فخر, فسق, طغي, عدو, شطط, سيئ, مسيء	Way	نجد, سبل, نساك, وسط, سنن, صرط, ضلك, طرق, سوي, نهج, داب, قوم, في, رشد, قرب, صرف, اله, سبب, رع, خرج, عرج, حكم, شكل, عرف, مثل	نجد, سبل, نساك, وسط, سنن, صراط, ضلك, طرق, سوي, طريق, سوي, منهج, داب, قوم, رشد, قرب, سبب, عرج, شكل, عرف
News	خير, نبا, بشر, ريب, اله, حدث, بشر, نبا, خير	خبر, نبا, بشر, نبأ, بشر, خير, نبأ, بشر	Fruit	زيت, نقر, قتل, حبيب, خردل, زرع, قطف, ثمر, اكل, فكه, طلع, جني, صرم	زيت, خردل, زرع, قطف, ثمر, اكل, فكه, طلع, جني, صرم
Life	برأ, بشر, انس, اول, حيي, عمر, نفس, دنو, بدو, عيش, هزز	برأ, انس, اول, حيي, عمر, نفس, دنو, بدو, عيش, هزز	Party	نفر, حزب, فءاي, طوف, فرق, شبع	نفر, حزب, فئاة, طائفة, فرق, شبيعة
Water	غسل, موه, سقي, حمم, ورد, عين	غسل, ماء, سقي, حميم, ورد, عين, تمع	Force	غضب, جند, كره, فاي, جمع, قوي	غضب, جند, فئاة, جمع, قوي
City	أفك, مصر, بلد, قري, مدن	مصر, بلد, قرية, مدن, مدينة	Neck	جيد, رقاب, طوق, معطف, عنق, حناجر	جيد, رقاب, طوق, معطف, عنق, حناجر
Reward	اله, حسن, اجر, ثوب, جزى, زيد	اله, حسن, اجر, ثوب, جزاء, زيد	Hear	سمع, شعر, انس, صمم, سمع, موت	سمع, علم, صمم, سمع, علم
Child	يتم, غلم, طفل, بني, ولد, نزر, كلل, صلح, صبو	يتم, غلام, طفل, بني, ولد, نزية, كلالة, صلح, صبوة, ابن	Kindness	من, معن, راف, حسن, عرف	من, معن, راف, حسن, عرف
Abode	دور, بيت, سكن, اوب, امم, ثوي, اوي, قرر	الدور, بيتك, سكن, اوب, ثوي, اوي, قرار, ماوي	Escape	فرر, فوز, حيص, عجز, عزب, صرف, وال, فوت, نوص, مناص, هرب	فرر, فوز, محيص, معجز, يعزب, صرف, فوت, نوص, مناص, هرب
Soul	قدس, روح, علي, نفس, وفي, عقب, روح	مقدس, علي, نفس, وفي, عقب, روح, بلطن	Determination	أفك, حرد, امر, اول, عزم, قدر	أفك, حرد, عزم, قدر
Sign	يقن, ابي, بين, اله, بصر	موقن, آية, بين, بصر, إشارة, رمز	Sacrifice	نحر, نساك, قرب, نصب, حلل, ذبح	أفك, نساك, قرب, نصب, حلل, ذبح
Desire	فاد, كون, هوي, رود, مني, همم, شهو, غير, اول, ارب, رجو, طمع, رود, رود, مني, اله	أفئدة, تكون, أهواء, آزاد, مني, هممة, شهوة, غير, أرجوا, طمع, حرص, حاجة, يشاء	Land	برر, حرث, بلد, ارض	برر, حرث, بلد, ارض
Strength	قوي, صبر, يدي, قوي, طوق, جبر, روح, محل, ازر, شدد, اول, ذا, ايد	صبر, يدي, قوي, طائفة, جبار, روح, ازر, شدد, ايد, بطش, شوكة	Blazing	شهب, سعر, شهب, مرج, شوط, لهب, لهب, جحم, سعر, حرق, لطي	شهب, سعر, شهب, مرج, شوط, لهب, لهب, جحم, سعر, حرق, لطي
Love	بتل, عكف, ودد, حبيب	عاكف, وودد, حبيب, مودة	Wealth	ترف, خير, مول, قنطر, ورث, طول, اله	ترف, خير, أموال, قنطر, ورث, طول, اله, رزق
Permission	رضو, انس, انن, رود, اله	رضوان, انس, انن, آزاد	Commit	جرح, اتى, غلو, قرف, سرف, زني, عتو	أتى, غلو, اقرت, اسرف, زان, تعنى
Harm	قرح, اذي, ضرر, غسل, قرح, جرح, ضرر, سوا, فتن, اذي, صير, عر	قرح, اذي, ضرر, ساء, فتن, اذي, صير, معزة, مصيبة			
Possessor	مرأ, نفس, عمر, فسق, سكن, ذو, ذا, اول, قوي, ايد, رفع, درج, باس, مرر, فضل, فرد	فرد, المرء, نفس, عمارة, فاسق, ذو, ذا, اول, رفع, درجة, مالك			

Table 8. System Result Not Valid Based Gold Standard

Word Input	System Result Not Valid
Veil	-
Name	قبس
Believe	قبل
Punishment	اتم, صنع, الم, مثل, نكر, ذا, لزم
Fire	وقع
Darkness	حما, دهم
Witness	قول, خوض, راعي

Word Input	System Result Not Valid
Stone	
Opinion	بَدُو, يَوْم, إِلَه, نَيْط
Book	أُمَّم, أَهْل
Story	-
Sound	سَلْم, اصْوَار
Rain	صَوْب
Animal	مَوْت, ذَا, سَرَق, عَيْد
Journey	كَيْز
Measure	كَان
Friend	قَاصِرَة, طَرْف, عَيْن
Owner	طُول
Protector	-
Evil	-
News	رَبب, إِلَه
Life	دُنُو, بَشَر
Water	-
City	إِفْك
Reward	-
Child	-
Abode	أُمَّم
Soul	-
Sign	غَيْر, أَوْل, إِلَه
Desire	غَيْر, أَوْل, إِلَه
Strength	أَوْل, ذَا, مَحَل
Love	بِتْوَل
Permission	إِلَه
Harm	غَسَل
Possessor	قَوِي, أَيْد, بَأْس, مَرَز, فَضَّل, سَكَن
Helper	ظَهَر, ظَهَر
Way	فِي, صَرَف, إِلَه, شَرَع, مَثَل, حَكَم
Fruit	صَلَام, نَقَر, قَيْبِل, حَنْب
Party	شَع
Force	كَرَة
Neck	-
Hear	مَوْت, شَعْر
Kindness	-
Escape	وَال
Determination	أَمْر, أَوْل
Sacrifice	-
Land	بِرَر
Blazing	-
Wealth	-
Commit	جَرَح

V. CONCLUSION

In this study the results of combining words that have the same meaning closeness are processed using Hierarchical Clustering and then processed by bending words that are close in meaning to words whose closeness means far from word entry. The results of this study can be used if all Muslims who are studying the book of the Qur'an and can add resources to Al-Qur'an research. Suggestions for further research are expected to develop word classes other than nouns. It is also hoped that further research can improve the accuracy of this assessment development.

APPENDIX

A. Gold Standard Result

Here is a table of results from the gold standard process based on some dictionary references obtained.

Table 9. Pure Gold Standard Result

Gold Standard Result	Translate
عِطَاء	Tabir, Penutup, Selubung
إِسْم	Nama, Kata benda
تَسْمِيَةٌ	Nama, Pemberian Nama
إِعْتِقَاد	Percaya, Ikatan
قِصَاصٌ	Balasan, Qishash
نَارٌ	Api
عَتَمَةٌ	Kegelapan
بَيِّنَةٌ	Bukti Nyata, Keterangan
بُرْهَانٌ	Bukti, petunjuk
دَلِيلٌ	Bukti, Tanda, Dalil
حَجَرِيٌّ	Batu, Berbatu
فِكْرٌ	Pikiran, Berfikir
ذَقْرٌ	Buku, Buku Catatan
حِكَايَةٌ	Kisah, Cerita, Hikayat
رَوَايَةٌ	Cerita, Narasi, Riwayat
نَعْمٌ	Nada, Melodi, Suara, Bunyi
غَيْثٌ	Hujan
حَيَوَانٌ	Binatang, Binatang Buas
رَحْلَةٌ	Perjalanan, Penjelajahan
حَجْمٌ	Volume, Ukuran, jarak
قِيَاسٌ	Takaran, Ukuran, Dimensi
ذُو	Mempunyai, Memiliki
مُنَاصِرٌ	Menolongnya, Penolong
سَيِّئٌ	Jahat, Jelek, Berbahaya
مُسِيءٌ	Jahat, Buruk, Jelek, Kejam
أَبَدٌ	Selama lamanya, Kekal, Abadi
نَمْعٌ	Air Mata, Berair
مَدِينَةٌ	Kota
إِبْنٌ	Anaknya, Anak Laki Laki

Gold Standard Result	Translate
مَأْوَى	Tempat tinggal, Kediaman
بَاطِن	Batin
حِرْص	Menginginkan, Kegemaran
حَاجَةٌ	Keperluan, Hajat, Kebutuhan
يَشَاءُ	Kehendaki, Keinginan
بَطْشٌ	Kekuatan, Sergapan
شَوْكَةٌ	Kekuatan, kekuasaan
مَوَدَّةٌ	Persahabatan, Cinta Kasih
مُصِيبَةٌ	Musibah, Bencana
مَالِكٌ	Raja, Merajai, pemilik
ظَهِيرٌ	Bantuan, Pembantu, Penolong
مُعَاوَنٌ	Tolong Menolong, Asisten
نِتَاجٌ	Produk, Buah Buahan, Panen
إِشَارَةٌ	Menunjuk, Isyarat, Sinyal
رَمْزٌ	Isyarat, Simbol, Tanda
شَيْعَةٌ	Golongan, Kelompok, Sekte
حَتَّاجِرٌ	Kerongkongan, Pangkal Tenggorokan
عَلِمَ	Mengetahui
إِحْسَانٌ	Kebaikan, Kedermawanan
هَرَبٌ	Lari, Menyelamatkan Diri
مُتَوَفِّدٌ	Menyala, Dinyalakan
رِزْقٌ	Rezeki

Word Input	System Result Valid by Gold Standard
Journey	يَضْرِبُونَ, سَيَّرَ, يَضْرِبُونَ, طَعَنَ, سَيَّرَ, بِالْحَجِّ, رَحَلَ, سَفَرَ, سَيَّرَ, رَحَلَ
Measure	وَطَأَ, حَسَبَ, حَسَابًا, كَيْلًا, قَدَّرَ
Friend	كَوَاعِبُ, صَاحِبُ, قَرْنٌ, رَفِيقٌ, حُورٌ, أَوْلِيَاءُ, خَلِيلٌ, عَشِيرٌ, صَاحِقٌ, حَمِيمٌ
Owner	عَرِشٌ, أَوْلَى, مَلِكٌ, أَهْلٌ, ذَا, نَقَمٌ
Protector	وَأَلِيهِ هَيْمَنٌ, حَفِظَ, وَكَّلَ, قُوَّتٌ, حِرْزَةٌ, صَاحِبٌ, حَفِظَ, وَأَلِيهِ قَوْمٌ, عَصَمَ, أَيْقَمَ, قَوْمٌ, وَكَّلَ, عَصَمَ, وَقِيَ, بَلَّغَ, حَفِظَ, حَرَسَ, رَصَدَ
Evil	سَوَاءَةٌ, بَأْسٌ, خَبِيثٌ, طَائِرٌ, رَجَزٌ, رَجِسٌ, أَشْرَارٌ, لُكْرٌ, غَوِيٌّ, هَجْرٌ, تُعْتَوَى, ذَنْبٌ, جُرْمٌ, فَجْرٌ, فِسْقٌ, طَغَى, عَدُوٌّ, شَطَطٌ
News	خَبْرٌ, خَبَرٌ, نَبَأٌ, بَشَرٌ, خَيْرٌ, نَبَأٌ, يَشْتَرُ
Life	بِرَاءٌ, أُنْسٌ, أَوْلَى, حَيِيَّةٌ, عُمُرٌ, نَفْسٌ, يُبْدِي, عَيْشٌ, هَرَاةٌ
Water	غَسَلًا, مَاءٌ, سَقِيَ, حَمِيمٌ, وَرَدًا, عَيْنٌ
City	مِصْرٌ, بَلَدٌ, قَرْيَةٌ, مَدَنٌ
Reward	إِلَهُ, حُسْنٌ, أَجْرٌ, ثَوْبٌ, جَزَاءٌ, زَيْدٌ
Child	بَيْتَمٌ, غِلَامٌ, طِفْلٌ, بَنِيهِ, وِلْدَانٌ, ذُرِّيَّةٌ, كِلَالَةٌ, صَالِحٌ, صَبِيحَةٌ
Abode	الدَّوْرُ, بَيْتِكَ, سَكَنَ, أَوْبٌ, مَثْوَى, أَوْيَ, قَرَارٌ
Soul	مُقَدَّسٌ, عَلَيَّ, نَفْسٌ, وَفِي عَقِبِ رُوحٌ
Sign	مُوقِنٌ, آيَةٌ, بَيِّنٌ, بَصَرَ, إِشَارَةٌ, رَمْزٌ
Desire	أَفِيدَةٌ, تُكُونُ, أَهْوَاءٌ, أَرَادَ, مَنِيَّةٌ, هَمَّةٌ, شَهْوَةٌ, غَيْرٌ, أَرْجَاوًا, طَمَعٌ, حِرْصٌ, حَاجَةٌ, يَشَاءُ
Strength	صَبْرٌ, يَدَيَّ, قُوَّةٌ, طَاقَةٌ, جَبَّارٌ, رُوحٌ, أَرْزُقُ, شَدِيدٌ, أَيْدٍ
Love	عَاكِفٌ, وَدُودٌ, حَبِيبٌ
Permission	رِضْوَانٌ, إِنْسَ, إِذْنٌ, أَرَادَ
Harm	فَرْحٌ, جُرْحٌ, ضَرَرٌ, سَاءٌ, فَنَنْ, أَذِي, ضَيْرٌ, مَعْرَةٌ
Possessor	فَرْدٌ, الْمَرْءُ, نَفْسٌ, عِمَارَةٌ, فَاسِقٌ, ذُو, ذَا, أَوْلَى, رَفَعٌ, نَرَجَّةٌ
Helper	ثَبِيحَةٌ, عَمِلَ, وَزِيرٌ, رَدَّ, نَصَرَ, مُصْرَخٌ, عَضُدٌ
Way	تَجَدُّ, سَبِيلٌ, لُسْكٌ, وَسْطٌ, سُنَنٌ, صِرَاطٌ, حَسْبٌ, طَرِيقٌ, سَوِيٌّ, مِثْمَاجٌ, ذَابٌ, قَوْمٌ, رُسُدٌ, قَرَّبٌ, سَبِيبٌ, خَرَجَ, عَرَجَ, شَكَلٌ, عُرْفٌ
Fruit	رَيْثٌ, خَزْدَلٌ, زُرْعٌ, فَطُوفٌ, ثَمَرٌ, أَكَلٌ, فَالِكِيَّةٌ, طَلَعٌ, حَنِيٌّ
Party	نَفَرٌ, حَرْبٌ, فِتْنَةٌ, طَائِفَةٌ, فِرْقٌ
Force	عَصَبٌ, جُنْدٌ, فِتْنَةٌ, جَمْعٌ, قُوَّةٌ
Neck	جِيدٌ, رِقَابٌ, طَوْقٌ, مَعْطَفٌ, عُنُقٌ
Hear	سَمِعَ, عَانَسَ, صَمَمَ, سَمِعَ
Kindness	مَتْنٌ, أَمْعَنَ, رُؤْفًا, حَسَنًا, عُرْفٌ
Escape	فَرَرٌ, فَوْزٌ, مَجِيصٌ, مُعْجَزٌ, يُعْرَبُ, صَرَفٌ, قُوَّتٌ, مَنَاصٌ
Determination	إِفْكٌ, حَرْدٌ, عَزْمٌ, قَدَّرَ
Sacrifice	أَنْحَرَ, لُسْكٌ, قَرَّبٌ, نَصَبٌ, حَلَلٌ, دُبْحٌ
Land	حَرْثٌ, بَلَدٌ, أَرْضٌ
Blazing	شِهَابٌ, سَعْرٌ, شِهَابٌ, مَارِحٌ, شَوَاطِئٌ, لَهَبٌ, لَهَبٌ, جَحِيمٌ, سَعْرٌ, حَرَقَ, تَلَطَّى
Wealth	ثَرَفٌ, خَيْرٌ, أَمْوَالٌ, قِنطَارٌ, وَرَثٌ, طَوْلٌ, إِلَهٌ

B. Valid System Result

Here is a table of valid system results based on the gold standard.

Table 10. Valid System Result

Word Input	System Result Valid by Gold Standard
Veil	لَبَسَ, ثَوْبٌ, سَرَبَلٌ, جَلَبٌ, حِجَابٌ, عَشَاوَةٌ
Name	وَسَمَّ, سَمُوًّا, إِلَهٌ, حَسَنٌ
Believe	صَدَّقَ, أَمَنَ, ظَنَّ
Punishment	عَذَابٌ, رَجَزٌ, نَكَالٌ, عُقُبٌ, بَأْسٌ, رَجَسٌ, حِرْصًا, عِقَابٌ
Fire	شَهْبٌ, نُورٌ, جَحِيمٌ, حَرَقَ, سَعْرٌ, سَمَمٌ
Darkness	سَوَادٌ, أَحْوَى, ظَلَمَةٌ, لَيْلٌ, غَسَقٌ, قَتَرٌ
Witness	نَظَرَ, رَقَبَ, بَصُرَ, شَاهَدَ, وَكَّلَ
Stone	صَخْرٌ, صَفْوَانٌ, حَجَرٌ, لُصْبٌ, حَصَبٌ
Opinion	دِينٌ, قَوْمٌ, حَكْمٌ, فَتْحٌ, فَصَلٌ, رَأْيٌ
Book	سِفْرٌ, كِتَابٌ, زُبُرٌ
Story	حَدِيثٌ, حَدِيثٌ, لَهْوٌ, سَطَرَ, حَدِيثٌ, فَصَّصَ, حَدِيثٌ, سَطَرَ, فَصَّصَ, نَبَأٌ
Sound	حَقَّقَتْ, صَوْتٌ, خَوَارٌ, صَدَقَ, سَوِيٌّ, رَجَزٌ, هَمْسٌ, حَسِييسٌ, صَوْتٌ
Rain	سَيْلٌ, طَلٌّ, عَصْرٌ, وَابِلٌ, مَطَرٌ, سَمَاءٌ, غَيْثٌ, وَدَقٌ, مَرُنٌ
Animal	أَنْتَعَى, تَنَمَّى, نَارِيٌّ, دَابَّةٌ, لِلْأَنَامِ, نَهِيمَةٌ, وَحْشٌ, هَدْيٌ, سَنَعٌ, جَارِحٌ, كَلْبٌ, دَابَّةٌ, نَكَرٌ

Word Input	System Result Valid by Gold Standard
Commit	أَتَى غَلَوُ، اِفْتَرَفَ، اَسْرَفَ، زَانَ، نَعَى

C. Accuracy Calculation Process

The following is the process of calculating system accuracy with a gold standard that uses Recall, Precision and F-Measure. Below this is the calculation used to calculate accuracy in this study. There are 364 elements for Nall, Nsystem is 454 elements, and 51 elements for Ngoldstandard.

In this calculation the details of the number of elements have been attached. NgoldStandar results are 51 elements contained in the appendix section Gold standard results, Nall results are valid system results based on the gold standard contained in the valid system results table in the appendix, NSystem results are system results that are attached to the test results. The following is a calculation method to produce accuracy from this research.

$$\text{Recall} : \frac{364}{454} = 80.1 \% \quad (5)$$

$$\text{Precision} : \frac{364}{51} = 71.4 \% \quad (6)$$

$$\text{FMeasure} : 2X \frac{80,1 * 71,4}{80,1 + 71,4} = 76 \% \quad (7)$$

ACKNOWLEDGMENT

I want to thank the lecturer who guided me until I finished this final project. I also thank my friends for the support provided, accompanying me to work late into the night. I thank my parents for giving me their prayers, guidance and support to do this final project or research until finished.

REFERENCES

- [1] H. A. R. El-Mazni, Pengantar Studi Ilmu Al-Qur'an Edisi 13, vol. 13, A. Zulfidar, Ed., Pustaka Al-Kautsar, 2015.
- [2] A. Y. Ali, "Quran by A. Yusuf Ali," [Online]. Available: theholyyquran.org.
- [3] "A Lexical Database for English," 2019. [Online]. Available: english.wordnet.princeton.edu.
- [4] "WordNet: A Lexical Database for English," in *Communications of the ACM* November 1995, No.11, Vol 38, 1995.
- [5] A. B. Dendy Sugono, *Tesaurus Bahasa Indonesia* Pusat Bahasa, Departemen Pendidikan Nasional, 2008.
- [6] T. K, *Tesaurus Bioteknologi : Sebagai Alat Bantu Pengindeksan Dokumen*, Pustakawan Madya PDII-LIPI.
- [7] A. Kigarriff, "Gold Standard Datasets for Evaluation Word Senses Disambiguation Program," University of Brighton, pp. -, Oktober 1998.
- [8] M. H. G. I Ketut Eddy Purnama, "Supervised Learning Indonesian Gloss Acquisition," in *IAENG International Journal of Computer Science*, 2015.
- [9] Persamaan Kata, "Indonesian Thesaurus," 2012 - 2020. [Online]. Available: Persamaankata.com.
- [10] M. A. B. Arini Rohmawati K, "Analysis of the commutative method approach on English thesaurus for developing synonym sets," *Indonesian Journal of Computing*, pp. 1-10, 2019.
- [11] M. A. B. A. F. H. Laras Gupitasari, "Pembangunan Sinonim Set Kosakata Al-Qur'an Dengan Pendekatan WordNet," *Jurnal Teknik Informatika dan Sistem Informasi*, pp. -, Maret 2020.
- [12] C. Z. Charu C. Aggarwal, *Mining Text Data*, Springer, 2012.
- [13] M. A. Bijaksana, "Slide PPT: CSH4H3 Text Mining Telkom University," 2019.
- [14] Suyanto, *Machine Learning Tingkat Dasar & Lanjut*, Informatika Bandung, Indonesia, 2018.
- [15] Z. N. M. A. Dongshik Kang, "A New Hierarchical Clustering," in *ICIIBMS2015, Track 2: Artificial Intelligence, Robotics, And Human-Computer Interaction*, Okinawa, Japan, 2015.