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

Salma Fauziah\(^1\)*, Moch Arif Bijaksana\(^2\)

\(^1\)\(^2\)Faculty of Informatics, Telkom University, Bandung, Indonesia
\(^1\)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

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](https://doi.org/10.20895/infotel.v12i3.477)
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 tokenization 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']            | [أوَّل, Owner] |
| ['Judgment’,’Conclusion’] | [حكم, Judgment, [حكم, Conclusion] |
| ‘Opinion’            | [قَدْرٌ, 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.

E. Clustering

After the dataset 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]. 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.

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.

\[
\text{SynonymPath}(C1,C2) = \frac{1}{\text{path}(c1,c2)} \tag{1}
\]

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 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 >= Threshold 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 contains the comparison table of grouping synonym set.

Table 4. Comparison Table of Grouping Synonym Set

| Before Grouping | After Grouping |
|-----------------|---------------|
| Veil, Garment, | Garment, Veil |
| غَشَاوَة, حِجَاب | حِجَاب, غَشَاوَة |

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.

Recall = \( \frac{\text{Nall}}{\text{Ngold}} \)  \( (2) \)

Precision = \( \frac{\text{Nall}}{\text{NSystem}} \)  \( (3) \)

FMeasure = \( \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)  \( (4) \)

In (2), Nall is the number of words in the synonym set that are in the gold standard, while NSystem is the number of words in the grouping synonym set produced by the system. In (3), Ngold 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       | [إسْمُ لَبَس, مِرِيَل, جَلَب, جَلَب] |
| Name       | [إسْمُ لَبَس, مِرِيَل, جَلَب] |
| Believe    | [إسْمُ لَبَس, مِرِيَل] |
| Punishment | [إسْمُ لَبَس, مِرِيَل, جَلَب] |
| Fire       | [إسْمُ لَبَس, مِرِيَل, جَلَب] |

Jurnal Infotel Vol.12 No.3 August 2020
https://doi.org/10.20895/infotel.v12i3.477
| Word Entry | Output Grouping Synonym Set |
|------------|-----------------------------|
| Darkness   | [فيض, غمس, سحاب] |
| Witness    | [譬, مثل, مثل, مثل] |
| Stone      | [مساح، مصداق، حب] |
| Opinion    | [نقطة، نقطة، نقطة] |
| Book       | [مصطلح، مصطلح، مصطلح] |
| Story      | [حالة، حالة، حالة] |
| Sound      | [صوت، صوت، صوت] |
| Rain       | [طقس، طقس، طقس] |
| Animal     | [نوع، نوع، نوع] |
| Friend     | [صديق، صديق، صديق] |
| Owner      | [حول، حول، حول] |
| Protector  | [حوار، حوار، حوار] |
| Evil       | [خط، خط، خط] |
| News       | [خبر، خبر، خبر] |
| Life       | [حياة، حياة، حياة] |

| Word Entry | Output Grouping Synonym Set |
|------------|-----------------------------|
| Love       | [حب، حب، حب] |
| Permission | [إذن، إذن، إذن] |
| Harm       | [ضرر، ضرر، ضرر] |
| Possessor  | [أولي، أول، أول] |
| Helper     | [ساعد، ساعد، ساعد] |
| Way        | [طريق، طريق، طريق] |
| Fruit      | [فواكه، فواكه، فواكه] |
| Sign       | [علامة، علامة، علامة] |
| Party      | [حفل، حفل، حفل] |
| Force      | [قوة، قوة، قوة] |
| Neck       | [عنق، عنق، عنق] |
| Kindness   | [أعمال، أعمال، أعمال] |
| Escape     | [هروب، هروب، هروب] |
| Determination | [قرار، قرار، قرار] |
| Sacrifice  | [حبوب، حبوب، حبوب] |
| Land       | [أرض، أرض، أرض] |
| Blaze      | [فحم، فحم، فحم] |
| Wealth     | [ثروة، ثروة، ثروة] |
| Commit     | [جريمة، جريمة، جريمة] |

### 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 غشاوة. 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 Lams 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':

![Fig.3. Process Comparison Between Previous Articles](image)

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 Almama Arabic Dictionary and the Munjdi 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.

| Word Input | System Result | System Valid + Gold Standard |
|------------|----------------|-------------------------------|
| Veil       | عينه، شعر، رأى | عينه، شعر، رأى |
| Name       | يرب، بديع، رجل | يرب، بديع، رجل |
| Believe    | يمسي، يلمسي، صرخ | يمسي، يلمسي، صرخ |
| Punishment | ضرب، ضرب، ضرب | ضرب، ضرب، ضرب |
| Fire       | نار، نار، نار | نار، نار، نار |
| Darkness   | أضواء، أضواء، أضواء | أضواء، أضواء، أضواء |
| Witness    | ينظر، ينظر، ينظر | ينظر، ينظر، ينظر |
| Stone      | حجر، حجر، حجر | حجر، حجر، حجر |
| Opinion    | رأي، رأي، رأي | رأي، رأي، رأي |
| Book       | كتاب، كتاب، كتاب | كتاب، كتاب، كتاب |
| Story      | قصة، قصة، قصة | قصة، قصة، قصة |
| Sound      | صوت، صوت، صوت | صوت، صوت، صوت |
| Rain       | غשם، غשם، غשם | غשם، غשם، غשם |
| Animal     | حيوان، حيوان، حيوان | حيوان، حيوان، حيوان |
| Journey    | رحل، رحل، رحل | رحل، رحل، رحل |
| Measure    | حساب، حساب، حساب | حساب، حساب، حساب |
| Friend     | واقع، واقع، واقع | واقع، واقع، واقع |
| Owner      | ملك، ملك، ملك | ملك، ملك، ملك |
| Protector  | حماي، حماي، حماي | حماي، حماي، حماي |

Jurnal Infotel Vol.12 No.3 August 2020
https://doi.org/10.20895/infotel.v12i3.477
| Word Input | System Result | Sytem Valid + Gold Standard |
|------------|---------------|----------------------------|
| Help       | Helper        |                            |
| Evolve     | Development   | Grouping of Synonym Set   |
| Permission | The Qur'an    | Thesaurus Vocabulary of   |
| Love       |                | in English using Hierarchical Clustering Algorithm |
| Fruit      |                |                            |
| Party      |                |                            |
| Force      |                |                            |
| Heat       |                |                            |
| Kindness   |                |                            |
| Escape     |                |                            |
| Determination |            |                            |
| Sacrifice  |                |                            |
| Land       |                |                            |
| Blazing    |                |                            |
| Wealth     |                |                            |
| Commit     |                |                            |

Table 8. System Result Not Valid Based Gold Standard

| Word Input  | System Result Not Valid |
|-------------|-------------------------|
| Veil        |                         |
| Name        |                         |
| Believe     |                         |
| Punishment  |                         |
| Fire        |                         |
| Darkness    |                         |
| Witness     |                         |
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.

| Gold Standard Result | Translate |
|-----------------------|-----------|
| غِطَاء              | Tabir, Penutup, Selubung |
| إِسم              | Nama, Kata benda |
| نَار              | Api |
| عَتَمَة         | Kegelapan |
| دَلِيل                  | Bukti Nyata, Keterangan |
| حَجَرِي                  | Bukti Tanda, Dalil |
| فِكْر                  | Batu, Berbatu |
| دَفْتَر                | Pikiran, Berfikir |
| حِكَايَة              | Buku, Buku Catatan |
| رِوَايَة                  | Kisah, Cerita, Hikayat |
| نَغَم                | Cerita, Narasi, Riwayat |
| غَيْث                  | Nada, Melodi, Suara, Bunyi |
| حَيَوَان              | Hujan |
| رِحْلَة              | Binatang, Binatang Buas |
| حَيْثَ                  | Perjalanan, Penjelajahan |
| دَمْع                | Volume, Ukuran, jarak |
| مَدِيْنَة            | Takaran, Ukuran, Dimensi |
| إبْن                | Mempunyai, Memiliki |
| سَيِّئ              | Menolongnya, Penolong |
| مُسِيء              | Jatuh, Jelek, Berbahaya |
| أَبَد                  | Jatuh, Buruk, Jelek, Kejam |
| حَبَّب                | Selama lamanya, Kekal, Abadi |
| حَيْثَ              | Air Mata, Berair |
| حَمْس                | Kota |
| نَمْع                | Anaknya, Anak Laki Laki |
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       | Journey                             |
| Name       | Measure                             |
| Believe    | Friend                              |
| Punishment | Owner                               |
| Fire       | Protector                           |
| Darkness   | Evil                                |
| Witness    | News                                |
| Stone      | Life                                |
| Opinion    | Water                               |
| Book       | City                                |
| Story      | Reward                              |
| Sound      | Child                               |
| Rain       | Strength                            |
| Animal     | Abode                               |
|             | Soul                                |
|             | Sign                                |
|             | Desire                              |
|             | Strength                            |
|             | Love                                |
|             | Permission                          |
|             | Harm                                |
|             | Possessor                           |
|             | Helper                              |
|             | Way                                 |
|             | Fruit                               |
|             | Party                               |
|             | Force                               |
|             | Neck                                |
|             | Hear                                |
|             | Kindness                            |
|             | Escape                              |
|             | Determination                       |
|             | Sacrifice                           |
|             | Land                                |
|             | Blazing                             |
|             | Wealth                              |

Gold Standard Result | Translate |
|----------------------|-----------|
| مأوى                 | Tempat tinggal, Kediaman |
| شيطان               | Batin      |
| حُرس                 | Meninginkan, Kegemaran |
| حاحة                | Keperluhan, Hajat, Kebutuhan |
| نَشة                 | Kehendaki, Keinginan |
| نَبْط              | Kekuatan, Sargapan |
| شَكْكة               | Kekuatan, kekuasaan |
| مَوَة                 | Persahabatan, Cinta Kasih |
| مَمْسِيبة             | Musibah, Bencana |
| مَالك                 | Raja, Merajai, pemilik |
| مَهْيَر               | Bantuan, pembantu, Penolong |
| مَخاوَن              | Tolong Menolong, Asisten |
| نَفَج                 | Produk, Bush Buhah, Panen |
| إِشْتَقاء             | Menunjuk, Isyarat, Sinyal |
| رَنْط               | Isyarat, Simbol, Tanda |
| شَيْعَة               | Golongan, Kelompok, Sekte |
| نَاجٍ                 | Kerongkongan, Pangkal Tengkorakan |
| عَمَل                 | Mengetahui |
| إِخْسَان             | Kebaikan, Kedemawan |
| هَرْب               | Lari, Menyelamatkan Diri |
| مَشْوَق             | Menyala, Dinyalakan |
| رُؤْك               | Rezeki     |

ISSN : 2085-3688; e-ISSN : 2460-0997
Development Grouping of Synonym Set Thesaurus Vocabulary of The Qur’an in English using Hierarchical Clustering Algorithm

Jurnal Infotel Vol.12 No.3 August 2020
https://doi.org/10.20895/infotel.v12i3.477
Development Grouping of Synonym Set Thesaurus Vocabulary of The Qur’an in English using Hierarchical Clustering Algorithm

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 N_all, N_system is 454 elements, and 51 elements for Ngoldstandard.

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

\[
\text{Recall} = \frac{364}{454} = 80.1 \%
\]
(5)

\[
\text{Precision} = \frac{364}{51} = 71.4 \%
\]
(6)

\[
\text{FMeasure} = \frac{2 \times \frac{80.1 \times 71.4}{80.1 + 71.4}}{80.1 + 71.4} = 76 \%
\]
(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: theholyquran.org.

[3] "A Lexical Database for English," 2019. [Online]. Available: english.wordnet.princeton.edu.