

# Semantic Text Relatedness On AlQur'an Translation Using Modified Path Based Method

(Case Study : From Surah Al-Ashr Until Surah An-Naas)

Yudi Irwanto<sup>1</sup>, M. Arif Bijaksana<sup>2</sup>, Adiwijaya<sup>3</sup>

School Of Computing, Telkom University, Indonesia<sup>1,2,3</sup>

yuditube@gmail.com<sup>1</sup>, arifbijaksana@telkomuniversity.ac.id<sup>2</sup>,  
adiwijaya@telkomuniversity.ac.id<sup>3</sup>

**Abstract.** Abdul Baquee Muhammad [1] have built Corpus that contained *AlQur'an* domain, *WordNet* and dictionary. He has did initialisation in the development of knowledges about *AlQur'an* and the knowledges about relatedness between texts in *AlQur'an*. The Path based measurement method that proposed by Liu, Zhou and Zheng [3] has never been used in the *AlQur'an* domain. By using *AlQur'an* translation dataset in this research, the path based measurement method proposed by Liu, Zhou and Zheng [3] will be used to test this method in *AlQur'an* domain to obtain similarity values and to measure its correlation value.

In this study the degree value is proposed to be used in modifying the path based method that proposed in previous research. Degree Value is the number of links that owned by a lcs (lowest common subsumer) node on a taxonomy. The links owned by a node on the taxonomy represent the semantic relationship that a node has in the taxonomy. By using degree value to modify the path-based method that proposed in previous research is expected that the correlation value obtained will increase.

After running some experiment by using proposed method, the *correlation* measurement value can obtain fairly good *correlation* ties with 200 Word Pairs derive from Noun *POS SimLex-999*. The *correlation* value that be obtained is 93.3% which means their bonds are pretty strong and they have very strong *correlation*. Whereas for the *POS* other than Noun *POS* vocabulary that owned by *WordNet* is incomplete therefore many pairs of words that the value of its similarity is zero so the correlation value is low.

**Keywords :** data mining, semantic, semantic text relatedness, semantic similarity, path based similarity, shortest path length, depth of subsumer, degree value.

## 1. Introduction

Semantic is the knowledge which learning about meaning of the word in the language. The semantic related with meaning of the word relation, it's like in synonym, antonym and hyponym. Morrish and Hirst 2004 [2] stated that semantic *relatedness* explains the power of semantic relationships between two words or concepts are measured. It includes various relationships among concepts, belonging classical relationships as hypernymy, hyponymy, meronymy, antonymy, synonymy, and any other 'non classical relations'.

Weeds [2] defined that semantic similarity is a specific case of *relatedness*, where the power of relatedness depends on the 'degree of synonyms', that is normally explained by classical relations. In the taxonomy or lexical hierarchy generally are shaped from hypernym and hyponym words. Hypernym is the words that have general meaning from the other words. Hyponyms are a group of

words that are of special significance meaning rather than measured words. Hyponym is a member of group of a hypernym meaning groups, eg hypernym = sport, hyponym = football, basketball, tennis, run.

Semantic similarity between word is often represented by similarity between concept that associated with the word. Concepts can have different meanings but remain semantically interconnected like sports and balls. Other instance that antonyms are respected has a semantic relationship such as smart and stupid, however they are dissimilar. A way that can be used to find out the value of similarity is to calculate the value of shortest path length between two words. The similarity value is also influenced by the depth of subsumer value of the lowest common subsumer (*lcs*) nodes from two concept. The similarity value measured between two concepts will increase if the depth of subsumer value of the two concepts also grows higher in the taxonomy.

## 2. Literature Review

### 2.1. Related Work

In the study that conducted by Liu, Zhou and Zheng [3] on a semantic similarity using path-based method managed to get a correlation value of 92.6%. The correlation value still has an opportunity to be improved because of the research by Liu, Zhou and Zheng [3] only optimizing the semantic relationship of hypernym and hyponym. In the Semantic relatedness there are many semantic relationships besides hypernym and hyponym such as synonym, antonym, meronym and holonym. Taking advantage of all semantic relationships is expected to increase the correlations value that have been achieved previously.

Abdul Baquee Muhammad [1] have built Corpus that contained *AlQur'an* domain, *WordNet* and dictionary. He has did initialisation in the development of knowledges about *AlQur'an* and the knowledges about the relatedness between texts in *AlQur'an*. The Path based measurement method that proposed by Liu, Zhou and Zheng [3] has never been used in the *AlQur'an* domain. In this research the path based measurement method proposed by Liu, Zhou and Zheng [3] will be used to test this method in *AlQur'an* domain to obtain similarity values and to measure its correlation value.

To obtain a better correlation value the degree value is proposed to be used in modifying the path based method that proposed by Liu, Zhou and Zheng [3]. Degree Value is the number of links that owned by a *lcs* (lowest common subsumer) node on a taxonomy. The links owned by a node on the taxonomy represent the semantic relationship that a node has in the taxonomy. By using degree value to modify the path-based method that proposed by Liu, Zhou and Zheng [3] is expected that the correlation value obtained will increase.

### 2.2. Semantic Similarity

Weeds [2] defined that semantic similarity is a specific case of *relatedness* where the power of relatedness depends on the 'degree of synonyms', that is normally explained by classical relations. Concepts that are semantically related are not necessarily similar such as car and fuel. Another example is that antonyms are considered to be semantically related such as beautiful and ugly, however they are dissimilar.

The semantic similarity between words has become a widespread topic in natural language processing and information retrieval for many years. It is widely used in a variety of applications such as sense disambiguation, detection and correction of malapropisms, multimodal documents retrieval, and automatic hypertext linking [3].

Similarity between words is often represented by similarity between concepts associated with the words. Many methods are developed to compute word similarity nowadays. Generally, these methods can be classified into two categories, one is edge counting based method that using only semantic link and second is corpus based methods that combining corpus statistics with taxonomic distance. The information sources that commonly be used in similarity measures are the shortest *path* length between words and the depth of subsumer in the taxonomy hierarchy [3].

The key of calculating semantic similarity lies in simulating human thinking behavior. Semantic similarity between words is decided by processing the first hand information sources in the human brain. Therefore the semantic similarity measurement can simulate human judgement by using multiple information sources [3].

Human tends to decide similarity between concepts through comparing their common features and different features. Therefore it be considered that the common and different features of two words to simulate the process of human judgement. It is assumed that human judgement for semantic similarity is a process of comparing equal attribute with aggregate attribute that is the sum of equal and different attribute [3]. The model of semantic similarity is shown in formula [1].

$$S(w_1, w_2) = \frac{f(d)}{f(d) + f(l)} \quad [1]$$

Where  $l$  denotes the shortest *path* length between  $w_1$  and  $w_2$ ,  $d$  denotes the depth of subsumer in the hierarchy semantic net,  $f$  is the transfer function for  $d$  and  $l$  [3].

Tabel 1 Reference Results

Semantic Similarity Method	Correlation with RG-28	Correlation with MC-28	Correlation with RG-65
WordNet Edges	0.740	0.739	0.787
Hirst-St. Onge	0.671	0.682	0.732
Jiang-Conrath	0.670	0.684	0.731
Leacock-Chodorow	0.801	0.820	0.852
Lin	0.773	0.814	0.834
Resnik	0.706	0.763	0.800
Yang-Powers	0.889	0.921	0.897
Li	0.891	0.883	0.881
Liu-Zhou-Zheng	0.909	0.926	0.891

Tabel 1 shows the performance of semantic similarity measures reported by previous researchers. From table 1 shows that Liu, Zhou and Zheng [3] methods achieves the best correlation 0.926 with the MC's average human value of 28 pairs of words compared with the whole method. This mean that Liu, Zhou and Zheng [3] method's is more accurate than most individual subjects. Therefore human judgement for semantic similarity can be simulated by the ratio of common features to the sum of common and different features between words.

### 3. Experiment Scenario

Figure 1 is a general stage of the system developed in this research. The system built in this research can perform the relatedness measurement and correlation measurement. The query is used as *WordNet* input to find their *Wordset*. In data preprocessing stage, the *Wordset* and 60 verses of *AlQur'an* are processed to make it being standart form. With Taxonomy and proposed method, the *Wordset* and 60 verses of *AlQur'an* are measured to find their relatedness score. With correlation method, the relatedness score and *SimLex* score are compared and measured to find their correlation score.

The general process of the system as follows

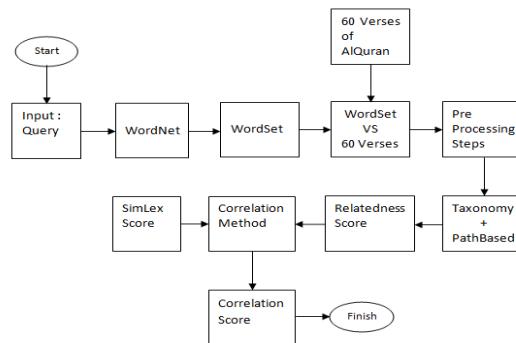


Figure 1 Experiment Scenario

### 3.1. Presentation of Wordset Dataset

Wordset is output from a query in the semantic sense of the word search process from the WordNet input, where the output shape set of semantic words and collected in one set of words. One word of WordNet input can produce multiple outputs, so the WordNet outputs are said can be collected in a set of words or Wordset. The query words that be used in this experiment are taken from the Pillar of Faith in Islam are faith, Allah, god, angel, holy, messenger, prophet, death, disaster, destiny.

Table 2 Query and Wordset

No	Query	Wordset
1	faith	[u'faith', u'organized_religion', u'religion', u'religious_belief', u'trust']
2	Allah	[u'Allah']
3	god	[u'God', u'Supreme_Being', u'deity', u'divinity', u'god', u'graven_image', u'idol', u'immortal']
4	angel	[u'Angel', u'Angel_Falls', u'angel', u'backer', u'holy_man', u'holy_person', u'saint']
5	holy	[u'holy', u'holy_place', u'sanctum', u'unholy']
6	messenger	[u'courier', u'messenger']
7	prophet	[u'oracle', u'prophesier', u'prophet', u'seer', u'vaticinator']
8	death	[u'Death', u'birth', u'death', u'deceive', u'demise', u'destruction', u'dying', u'end', u'expiry', u'last']
9	disaster	[u'calamity', u'cataclysm', u'catastrophe', u'disaster', u'tragedy']
10	destiny	[u'circumstances', u'destiny', u'fate', u'fortune', u'lot', u'luck', u'portion']

### 3.2. Presentation of 60 verses of AlQuran Dataset

The second dataset is a dataset that containing 60 verses translations of AlQur'an. The second dataset must be processed through preprocessing stage so the output can be more accurately. There are three steps in the preprocessing process, those process namely TSS (tokenization, stopword and stemming). 60 verses of AlQur'an translation comes from surah Al- 'Asr verse 1 to the surah of An-Naas verse 6.

### 3.3. Presentation of SimLex-999 Dataset

The correlation value calculation need dataset that be used to compare data from similarity measurement. The dataset can be extracted from SimLex-999 [13], while the SimLex-999 has 999 words pair which completed with their similarity score. The similaity measurement needs 200 words pair that be used in the computation. 200 words pair that taken from SimLex-999 are completed with their similarity score, then 200 words pair are used in similarity computation through proposed method.

### 3.4. The System Improvement

Keijo Ruohonen [16] defined that degree of the vertex is the number of edges with vertex as and end vertex. Number of edges that exist at a vertex are the edge that derived from that vertex or towards to the vertex. The degree value can be used to compute the number of the edge that established at the

vertex. Each vertex in the vertex network will have a degree because each vertex will have a relationship with another vertex.

*WordNet* has a word hierarchy or commonly called taxonomy. In the taxonomy there are relationship between *synsets*, the relationships are synonym, antonym, hypernym, hyponym, meronym and holonyms. That relationship between *synsets* are connected by several links to form a word or *synsets* network. Each *synset* in taxonomy can have multiple links depending on the relationships that are formed and owned with other *synsets*.

Therefore the degree value can be obtained from the number of links formed from a node connected to another node through the hypernym, hyponym, meronym and holonym relationships. And degree value which is proposed to modify the formula is the number of edge which exists at the lowest common subsumer (*lcs*) vertex. Therefore the degree value can be proposed to modify the formula derive from Liu, Zhou and Zheng [3], and the formula changed like shown below.

$$S(w_1, w_2) = \frac{b*d}{b*d + 1} \quad [2]$$

where *b* is the degree value from the lowest common subsumer (*lcs*) node.

In the *path based* method that proposed by Liu, Zhou, and Zheng [3] the *semantic* relationship that be used is the relationship of hypernym and hyponym only. While in *semantic* relatedness there are more existing relationships, not only hypernyms and hyponyms relationships but also synonyms, antonyms, meronyms and holonyms relationships. These relationships have not been optimized in the *path based* method that proposed by Liu, Zhou, and Zheng [3].

In the proposed method it uses the variable degree value, where the variable degree value is the number of links that owned by a node. Those links are representations of *semantic* relationships that owned by a node in *WordNet* taxonomy. So by using the variable degree value all existing semantical relationships can be accommodated well in the proposed method.

#### 4. Result

From the experiments that have been done, in accordance with the hypothesis, with the addition of degree value in *semantic relatedness* method has proven can improve the performance of relatedness value measurement. For analyzis the effectiveness of the proposed method, each *POS* in *WordNet* are measured their *correlation* value. The following are the result of *relatedness* value measurement and *correlation* value measurement.

##### 4.1. Relatedness Value Measurement

In the tables displayed in this research, Path denotes method used by Liu, Zhou and Zheng [3] in doing research that is path based method, Proposed denotes the method that be proposed in this research.

Table 3 Best *relatedness* value based on the query in AlQur'an domain

NO	Set of Query	Method		Best Relatednees Value of Verses of AlQur'an	
		Path	Proposed	Document	Verses of AlQuran
1	faith	0.664	0.862	d37	109:6 For you is your religion, and for me is my religion.
2	Allah	0.577	0.980	d37	109:6 For you is your religion, and for me is my religion.
3	god	0.427	0.683	d37	109:6 For you is your religion, and for me is my religion.
4	angel	0.309	0.702	d56	114:2 The Sovereign of mankind.
5	holy	0.170	0.693	d56	114:2 The Sovereign of mankind.
6	messenger	0.560	0.987	d56	114:2 The Sovereign of mankind.
7	prophet	0.503	0.984	d56	114:2 The Sovereign of mankind.
8	death	0.279	0.731	d1	103:1 By time,
9	disaster	0.326	0.747	d1	103:1 By time,
10	destiny	0.286	0.703	d1	103:1 By time,

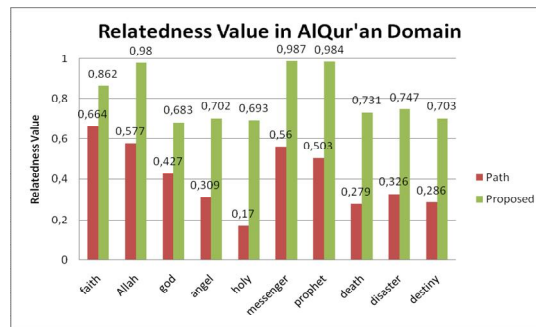


Figure 2 Relatedness Value in AlQur'an Domain

Table 4 Best *relatedness* value based on the query in General domain

NO	Set of Query	Method		Best Relatedness Value of Sentences of News	
		Path	Proposed	Document	Sentences
1	faith	0.149	0.508	d30	The sound of a trumpet blast
2	Allah	0.127	0.344	d29	The sun rises from the western horizon
3	god	0.172	0.484	d29	The sun rises from the western horizon
4	angel	0.268	0.513	d22	The rise of Imam Mahdi
5	holy	0.167	0.557	d29	The sun rises from the western horizon
6	messenger	0.484	0.733	d22	The rise of Imam Mahdi
7	prophet	0.437	0.726	d22	The rise of Imam Mahdi
8	death	0.145	0.513	d30	The sound of a trumpet blast
9	disaster	0.142	0.489	d30	The sound of a trumpet blast
10	destiny	0.159	0.519	d30	The sound of a trumpet blast

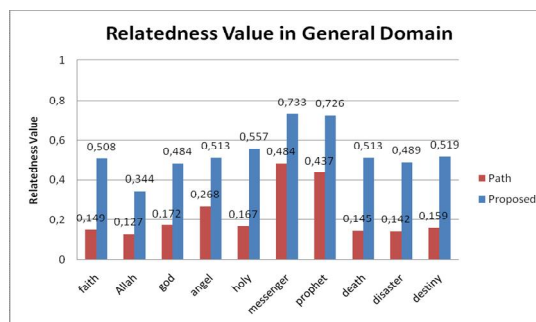


Figure 3 Relatedness Value in General Domain

This experiment is run to find out the highest relatedness value that can be obtained by each method on *AlQur'an* domain or general domain. At Figure 2 shows the result from the experimental measurements in *AlQur'an* domain indicate that the value of *relatedness* measurements by using the proposed method get better result than the Path based method proposed by Liu, Zhou and Zheng [3]. And at Figure 3 shows the result from the experimental measurements in General domain indicate that the value of *relatedness* measurements by using the proposed method get better result than the Path based method proposed by Liu, Zhou and Zheng [3]. From the experiment of *relatedness* measurements value in the both domains obtained results that the value of *relatedness* measurements by using the proposed method get highest result better than the Path based method proposed by Liu, Zhou and Zheng [3].

#### 4.2. Analyzis in Part of Speech of Wordnet

This experiment is running to measure the effectiveness and reliability of the method. The analyzis process that will be performed is comparing the results output from *relatedness* calculation process using the proposed method and with a list of *similarity* score of words pairs that come from *SimLex-999*. Both values are compared to check their *correlation* whether the change in value on the results of



the calculation process in line with changes in the *similarity* score of the pair of words from *SimLex-999* or in the reverse line or completely unrelated to the *Pearson* method.

The details about strength of *correlation* has explained by Jonathan Sarwono [15]. The strength of *correlation* according to Jonathan Sarwono [15] are shown at Table 5.

Table 5 Degree of *Correlation* by Jonathan Sarwono [15]

No	Range	Degree of Correlation
1	0	no correlation
2	0.001 - 0.250	the correlation is very weak
3	0.251 - 0.500	correlation enough
4	0.501 - 0.750	strong correlation
5	0.751 - 0.999	the correlation is very strong
6	1	perfect correlation

After through some *correlation* value computation step which involve some *POS dataset*, the *correlation* value from three *POS dataset* will be displayed and be compared each other. The comparison is performed to find where the *POS* parameter that can make *correlation* value obtain maximal result or to find where the *POS* parameter that cannot make *correlation* value obtain maximal result. The *correlation* value which computed with three *POS* parameter *dataset* are shown at Table 6 and Figure 4.

Table 6 Comparison *correlation* value with three *POS dataset*

POS	Maximum Correlation Value		Degree of Correlation
	Path	Proposed	
Adjective POS	0.101	0.015	the correlation is very weak
Noun POS	0.780	0.933	the correlation is very strong
Verb POS	0.339	0.279	correlation enough

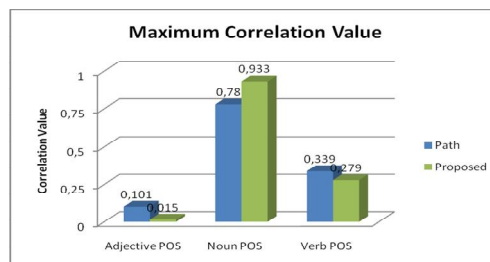


Figure 4 *Correlation* Value Diagram

#### 4.3. Analisis in Variable of Proposed Method

Proposed methods consist of three variables, those variables are variable *d* denotes depth of subsumer, variable *l* denotes shortest path length between two concept and variable *b* denotes degree of lcs node or number of links in the lcs node. This experiment is running to analysis the variables to find the variable that most influence in the changing of similarity value and correlation value and the variables that are less influential on changes in similarity value and correlation value.

After through some similarity value and *correlation* value computation with involving the changing of varibale *d*, *l* and *b* which each variable has value 1, the similarity value and *correlation* value from three variables will be displayed and be compared each other. The comparison is performed to find where the variable that more influences in similarity value and *correlation* value measurement and to find where the variable that less influences in similarity value and *correlation* value measurement. The comparison of similarity value and correlation value are shown at Table 7 and Figure 5.

Table 7 Comparison of deviation value in relatedness value measurement

Deviation Value	Proposed d = 1	Proposed l = 1	Proposed b = 1
Relatedness Value	0.083	0.128	0.380

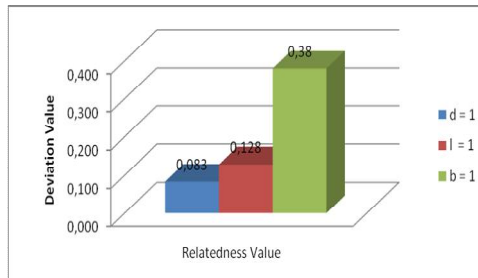


Figure 5 Comparison of deviation value in relatedness value measurement

Table 8 Comparison of deviation value and correlation value measurement

Deviation Value	Proposed d = 1	Proposed l = 1	Proposed b = 1
Correlation Value	0.027	0.267	0.146

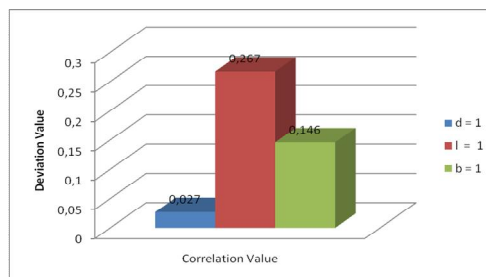


Figure 6 Comparison of deviation value in correlation value measurement

### 5. Discussion

From the result of the *relatedness* computation experiment revealed that proposed method obtain better *relatedness* value if be compared with *Path based* method that used in Liu, Zhou and Zheng [3] research. Proposed Method obtains better result because its involves degree value in the path based formula so it can increases *relatedness* value and correlation value.

Tthe degree value is used in this proposed method, degree value has means the number of links that owned by a node. Those links are representations of *semantic* relationships that owned by a node in *WordNet* taxonomy. So by using the variable degree value all existing semantical relationships can be accommodated well in the proposed method.

Based on the *correlation* computation experiment indicated that parameter which influences in *correlation* measurement is Noun *POS dataset*. With Noun *POS dataset*, *correlation* value obtains 78% using path method and 93,3% using proposed method. The result has mean that *relatedness* computation result has *correlation* with Noun *POS dataset* from *SimLex-999* and the strength of their *correlation* is very strong. The lowest *correlation* with *relatedness* computation result is adjective *POS dataset* from *SimLex-999* because their *correlation* value just 10,1% using path method and 1,5% using proposed method.

For Noun *POS* obtains high correlation value because in *WordNet* for Noun *POS* has a more complete vocabulary so rarely there are pairs of words whose value of similarity is zero so that the correlation value becomes high. Whereas for the *POS* other than Noun *POS* vocabulary owned



*WordNet* is incomplete so many pairs of words that the value of its similarity is zero so the correlation value is low.

From analysis of the proposed method and based on the experiment of the similarity value measurement, the most influential variable is the variable  $b$  which has deviation value 0.380 and the less influential is the variable  $d$  which has deviation value 0.083. In the measurement of correlation value, the most influential variable is the variable  $l$  which has deviation value 0.267 and the less influential is the variable  $d$  which has deviation value 0.027.

The variable  $b$  becomes the most influential variable in the relatedness value measurement because the variable  $b$  is the number of links owned by a node on taxonomy and a node in taxonomy *WordNet* has many semantic links so that when value  $b$  is made minimalis (1) it will be very influential to the relatedness value obtained. The variable  $l$  becomes the most influential variable on the correlation value measurement because the value of  $l$  is the distance or the number of links formed between two words (concept) measured on the taxonomy and the distance between the concept in taxonomy *WordNet* far apart so that when the value  $l$  is made minimalis (1) will be very influences the correlation value obtained.

## 6. Conclusions & Future Work

### 6.1. Conclusions

By involving the degree value in the *path based* formula, proposed method will get better *relatedness* value while in the example query 'messenger' and the verse is 114:2, it can reaches 98,7% compared by using the path based method that proposed by Liu, Zhou and Zheng [3] its reaches 56%.

With the proposed method, *correlation* value measurement can obtain fairly good *correlation* ties with 200 Word Pairs derive from Noun *POS SimLex-999*. 93.3% is the maximum *correlation* score that can be reached in this research, that value has means their bonds are pretty strong and they have very strong *correlation*.

The *correlation* value measurement in Adjective *POS* and Verb *POS* do not obtain high result as Noun *POS* result because for the *POS* other than Noun *POS* vocabulary that owned by *WordNet* is incomplete, therefore many pairs of words that the value of its similarity is zero so it makes the correlation value is low.

Measurement of correlation value is influenced by relatedness value that obtained from calculation process and relatedness value that obtained from external source such as *Simlex-999* which *Simlex-999* is used as human judgment and comparator in calculation of correlation value.

Based on the proposed method analysis in the similarity value measurement, the most influential variable is degree value which has deviation value 0.380 and the less influential is depth of subsumer which has deviation value 0.083. In the measurement of correlation value, the most influential variable is shortest path length which has deviation value 0.267 and the less influential is depth of subsumer which has deviation value 0.027.

This study has several side that still can be developed to achieve better value for *relatedness* and to assist the development of research on text and *AlQuran* both by the Qur'anic and text researchers or the academic community from all colleges.

### 6.2. Future Works

The number of existing methods in the measurement of text *relatedness* such vector based method, information content based method, gloss based method can be tried in the experiment to open opportunities for possible improvement that can be found in the text *relatedness* research in the future work later.

By using other knowledge bases besides the *Wordnet*, there is a big opportunity to get an increasing of *relatedness* value, because there are some limitation in this experiment so it has not had enough time to test. This would open the opportunities for future work in the development of research on the *semantic text relatedness* in the *AlQuran* domain.

With so many verses in the *AlQuran* which reached 6622 verses which in this research has not been performed, additional number of verses or use other set of verses in the *AlQuran* can be used in the subsequent research to open opportunities the possibility of increasing *relatedness* value that will be get.

Due to several limitations in this study led to the possibilities that be mentioned above have not had enough time to do the research to find the likely possibilities that can be obtained from several variations of the research and experiments that could unlock the barriers that may exist and hinder the increasing relatedness value.

## Bibliography

- [1] Abdul Baquee Muhammad, “ Annotation of Conceptual Co-Reference and Text Mining the Quran “, Submitted in accordance with the requirements for the degree of Doctor of Philosophy, England, September 2012.
- [2] Ziqi Zhang; Anna Lisa Gentile; Fabio Ciravegna, “Recent Advances in Methods of Lexical Semantic Relatedness - a Survey”, *Natural Language Engineering*, Volume 19, Issue 04, October 2013.
- [3] Xiao-Ying Liu; Yi-Ming Zhou; Rou-Shi Zheng, “Measuring Semantic Similarity in Wordnet”, *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, Hongkong, 19-22 August 2007.
- [4] Samer Hasan; Rada Mihalcea, “Semantic Relatedness Using Salient Semantic Analysis”, *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, Texas, 2011.
- [5] Abdul Baquee Muhammad Sharaf, Eric S.Atwell, “QurSim : A Corpus for Evaluation of Relatedness in Short Texts”, *Proceedings of the Eighth International Conference on Language Resource and Evaluation (LREC)*, Istanbul, 21-27 May 2012.
- [6] Akip Maulana, Moch. Arif Bijaksana, Ir., M.Tech, PhD, M. Syahrul Mubarak, ST., M.T, “Perancangan Semantic Similarity based on Word Thesaurus Menggunakan Pengukuran Omiotis Untuk Pencarian Aplikasi pada I-GRACIAS”, *Indonesian Symposium on Computing*, Sept 2016.
- [7] Lingling Meng, Runqing Huang, Junzhong Gu, “A Review of Semantic Similarity Measure in WordNet”, *International Journal of Hybrid Information Technology*, Vol. 6, No.1, January 2013.
- [8] Lingling Meng, Runqing Huang, Junzhong Gu, “Measuring Semantic Similarity of Word Pairs using Path and Information Content”, *International Journal of Future Generation Communication and Networking (IJFGCN)*, Vol. 7, No.3, 2014.
- [9] Hongzhe Liu, Pengfei Wang, “Assessing Text Semantic Similarity Using Ontology”, *Journal of Software*, Vol. 9, No.2, February 2014.
- [10] Issa Atoum, Ahmed Otoom, “Efficient Hybrid Semantic Text Similarity using Wordnet and a Corpus”, *International Journal of Advanced Computer Sciene and Applications*, Vol. 7, No.9, 2016.

- [11] Tingting Wei, Huiyou Chang, “Measuring Word Semantic Relatedness using WordNet-based Approach”, Journal of Computer, Vol. 10, No.4, July 2015.
- [12] Gang Liu, Ruili Wang, Jeremy Buckley, Helen Min Zhou, “A WordNet-based Semantic Similarity Measure Enhanced by Internet - based Knowledge”, Proceedings of the Twenty Third International Conference on Software Engineering & Knowledge Engineering (SEKE-2011), Miami, 7-9 Juli 2011.
- [13] Felix Hill, Roi Reichart, Anna Korhonen, “SimLex-999 : Evaluating Semantic Models with (Genuine) Similarity Estimation”, Journal of Computational Linguistics, Vol. 41, No.4, December 2015.
- [14] Fellbaum, Christine, “About Wordnet, A Lexical Database for English”, 2005, [Online], Available at : <https://wordnet.princeton.edu/>.
- [15] Jonathan Sarwono, “Mengenai Apa Itu Analisis Korelasi”, 2006, [Online], Available at : [http://www.jonathansarwono.info /](http://www.jonathansarwono.info/).
- [16] Keijo Ruohonen, “Graph Teory”, 2013.
- [17] Budi Raharjo, “Mudah Belajar Python untuk Aplikasi Desktop dan Web”, 2015.