

Development Synonym Set for the English Wordnet Using the Method of Comutative and Agglomerative Clustering

Munirsyah^{[1]*}, Moch. Arif Bijaksana^[2], Widi Astuti^[3]

Bachelor of Informatics Engineering, Faculty of Informatics
Telkom University, Bandung, Indonesia

^[1]munirsyahh@student.telkomuniversity.ac.id, ^[2]arifbijaksana@telkomuniversity.ac.id, ^[3]astutiwidi@telkomuniversity.ac.id

I. INTRODUCTION

The English Wordnet application has previously been developed by Princeton University in the United States which aims to model the lexical knowledge of native English speakers, where the results of its development take the form of desktop-based applications. Putting conventional dictionaries on line seems a simple and natural marriage of the old and the new [1].

Wordnet becomes a system that can provide information automatically because it has a dictionary concept by using a searching method rather than a dictionary in general that uses the alphabetical order [2].

The smallest unit of Wordnet is the Synonym set which represents the specific meaning of a word [3]. Each Synonym set contains the form of a Synonym set of words and a semantic pointer that explains the relationship between one Synonym set and another Synonym set [1]. Synonym set can also be called a collection of words in Wordnet that can represent one meaning, apart from the representation of meaning, a word also has a relationship between words such as hypernym, hyponym, holonym, meronym, and others [4].

In Wordnet development, synonyms of related or commutative words are needed, synonyms of words are obtained from English Thesaurus, the word has several synonym meanings and is used as data. To get commutative data, the data is processed using the commutative method, the commutative method is a data processing technique by comparing the first word with the second word, in the comparison the data checking process is related or commutative, said commutative if first word has the meaning of second word and second word has first word. The process of the commutative method produces the synonym set.

For the next process, a synonym set using Clustering, Clustering is used to group words that are similar. In this study the clustering used is Agglomerative Clustering. Agglomerative Clustering is a bottom-up Hierarchical Clustering method that combines n clusters into one single cluster [5].

According to tan Agglomerative clustering has the advantage that it does not need to determine the number of clusters and does not take into account initial centroids. these two things show the accuracy of using agglomerative clustering because to build wordnet data to be processed will be very much and not optimal if there is a cluster value in the grouping of data.

In the agglomerative clustering technique process there is a threshold value to determine the number of iterations or as a condition to stop the iteration process. The clustering process in this study will use a threshold value of 0.1 to 1 to test the best threshold value to produce the best Synonym set and calculate its accuracy value. Accuracy calculation and evaluation will use the F-measure method to find the best results [1] [5].