Synonym Sets Extraction Menggunakan Metode *Agglomerative Clustering* untuk Membangun WordNet Bahasa Indonesia

Bayu Setya Ajie Perdana Putra[1]**, Moch. Arif Bijaksana**[2]**, Diyas Puspandari**[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]bayusapp@students.telkomuniversity.ac.id, [2]arifbijaksana@telkomuniversity.ac.id,
[3]diyaspuspandari@telkomuniversity.ac.id

Abstract
Research and application development in the field of Natural Language Processing (NLP) and Computational Linguistics in Indonesian are still not widely found. This is caused by limitations in monolingual lexical resources in the Indonesian form. At present for lexical monolingual sources in the form of Indonesian, only available a Large Indonesian Dictionary and Indonesian Thesaurus issued by the Indonesian Language Center. To support this research, a lexical database of the Indonesian language is needed, which is to build WordNet in Indonesian. However, before building WordNet a basic component is needed which consists of a set of synonyms (synchronization) to determine the inter-word semantic relationship. This research will use synonym sets originating from the Big Indonesian Dictionary and the Indonesian Thesaurus which will then be extracted. The evaluation results of this study produce an F-measure or F1 score of 82% as a verification value generated by the clustering process and the results of the clustering process have been validated with linguists

Keywords: WordNet, Synonym Sets, Agglomerative Clustering.