

#### **Abstrak**

Akan dibangun sebuah kamus(*wordnet*) yang merupakan *lexical database* bahasa Indonesia. *Synset* merupakan bagian yang dibentuk pada tahap awal pembangunan *lexical database*. Pembangunan yang penulis kerjakan ini pada tahap kedua yaitu menggabungkan beberapa *synset* yang dipandang mirip.

*Synset* hasil ekstraksi Tesaurus Bahasa Indonesia akan digabungkan berdasarkan nilai *similarity* dan *distance value*. *Hierarchical clustering* sendiri adalah metode yang mengadaptasi teknik clustering. Alasan menggunakan teknik ini karena kesederhanaan dari teknik itu sendiri dan hasil dari kelompok bisa tidak diprediksi sebelum proses pengelompokan dilakukan.

Hasil utama dari tugas akhir ini adalah kumpulan *synonym set* yang sudah tercluster sesuai kemiripannya dan sudah dihitung nilai akurasi. Nilai perhitungan akurasi yang didapat yaitu sebesar 69,88%.

**Kata kunci:** *Wordnet, Hierarchical clustering, Tesaurus Bahasa Indonesia , Synset.*

---

#### **Abstract**

A dictionary will be built (*wordnet*) which is a lexical database in Indonesian. *Synset* is a part that was formed in the early stages of lexical database development. The development that the author is working on in the second stage is to combine several *synets* that are considered similar.

This topic is important because a dictionary will be built (*wordnet*) which is a lexical database in Indonesian. *Synset* is a part that was formed in the early stages of lexical database development. The development that the author is working on in the second stage is to combine several *synets* that are considered similar.

The *synset* of the Indonesian thesaurus extraction results will be combined based on similarity and distance value. *Hierarchical clustering* itself is a method that adapts clustering techniques. The reason for using this technique is because of the simplicity of the technique itself and the results of the group can not be predicted before the grouping process is done.

The main result of this final assignment is a set of *synonym sets* that have been clustered according to similarity and the accuracy value has been calculated. The calculated value of accuracy obtained is 69,88%.

**Keywords:** *Wordnet, Hierarchical clustering, Tesaurus Bahasa Indonesia , Synset.*