

Analisis Pembangunan *Word Sense* pada WordNet Bahasa Indonesia Menggunakan Metode *Hierarchical Clustering*

Laily Dwi Anggaraini¹, Moch. Arif Bijaksana², Diyas Puspandari³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹lailyangga@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

³diyaspuspandari@telkomuniversity.ac.id

Abstrak

Dalam pembangunan wordNet Bahasa Indonesia, *synonym set* merupakan salah satu bagian dari pembangunan wordNet yang merupakan kumpulan kata yang merepresentasikan suatu makna. *Synonym sets* yang dibangun bersumber dari Tesaurus Bahasa Indonesia yang diproduksi oleh Pusat Bahasa dengan proses ekstraksi dari Tesaurus Bahasa Indonesia tersebut didapatkan beberapa *synonym set* yang memiliki kemiripan/kesamaan makna (*word sense*) dimasing-masing kata. Perbedaan antara WordNet dengan kamus bahasa pada umumnya adalah kamus bahasa memfokuskan pada kata sedangkan WordNet memfokuskan diri kepada makna kata yang kemudian dilakukan proses *clustering* untuk menemukan *synonym set* yang lebih baik lagi. Oleh karena itu digunakan metode *Hierarchical Clustering* untuk mengelompokkan *synonym set* yang mirip berdasarkan nilai *similarity* dan *distance value*. *Synonym set* yang telah melalui proses *clustering* tersebut dapat digunakan sebagai salah satu pembangunan *lexical database*. Hasil utama dari penelitian ini adalah kumpulan *synonym set* yang sudah melalui proses *clustering* dan sudah dihitung nilai akurasinya.

Kata kunci : *WordNet, Hierarchical Clustering, Tesaurus Bahasa Indonesia, Word Sense*