

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

Bayu Setya Ajie Perdana Putra¹, Moch. Arif Bijaksana², Diyas Puspandari³

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

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

³diyaspuspandari@telkomuniversity.ac.id

Abstrak

Penelitian dan pengembangan aplikasi di bidang *Natural Language Processing (NLP)* dan *Computational Linguistics* dalam Bahasa Indonesia masih belum banyak ditemukan. Hal ini disebabkan oleh adanya keterbatasan dalam *monolingual lexical resource* dalam bentuk Bahasa Indonesia. Saat ini untuk *monolingual lexical resource* dalam bentuk Bahasa Indonesia hanya terdapat Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia yang dikeluarkan oleh Pusat Bahasa Indonesia. Untuk mendukung dalam penelitian tersebut diperlukan suatu *lexical database* Bahasa Indonesia yaitu membangun WordNet Bahasa Indonesia. Namun, sebelum membangun WordNet diperlukan komponen dasar berupa *synonym sets (synsets)* untuk menentukan hubungan simantik antarkata. Dalam penelitian ini akan digunakan *synonym sets* yang berasal dari Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia yang selanjutnya akan dilakukan proses ekstraksi *synonym sets* menggunakan metode *agglomerative clustering* untuk mengelompokkan *synonym sets* menjadi lebih optimal dan mempermudah dalam proses pembangunan WordNet Bahasa Indonesia. Hasil evaluasi dari penelitian ini menghasilkan nilai *F-measure* atau *F1 score* sebesar 82% sebagai nilai akurasi yang dihasilkan oleh proses *clustering* dan hasil dari proses *clustering* tersebut telah divalidasi bersama ahli bahasa.

Kata kunci : WordNet, *Synonym Sets*, *Agglomerative Clustering*