

1. Pendahuluan

Bagian ini menjelaskan tentang latar belakang, topik dan batasan, serta tujuan dari penelitian tugas akhir.

Latar Belakang

Princeton WordNet (PWN) adalah suatu *lexical database* dalam bentuk WordNet Bahasa Inggris yang dikembangkan oleh Princeton University dan dibangun secara manual oleh sekelompok *lexicographers* [1] [10]. PWN berisi *database* kelas kata, yaitu kata benda, kata kerja, kata sifat, dan kata keterangan yang dikelompokkan ke dalam *synonym sets (synsets)* [9]. PWN dapat digunakan untuk penelitian dan pengembangan aplikasi dalam bidang linguistik, *Natural Language Processing (NLP)*, *Text Mining*, dll [1]. Karena hal inilah banyak dibuat dan dikembangkan WordNet dalam bentuk bahasa lain [9] [10].

Namun, dalam Bahasa Indonesia penelitian terkait hal tersebut masih belum banyak dilakukan karena adanya keterbatasan *lexical database* dalam bentuk Bahasa Indonesia yang dapat digunakan secara bebas. Untuk saat ini, *monolingual lexical resource* dalam bentuk Bahasa Indonesia hanya ada dalam bentuk Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia yang dikeluarkan oleh Pusat Bahasa Indonesia dan dapat digunakan secara bebas. Edisi terbaru dari Kamus Besar Bahasa Indonesia yang dikeluarkan oleh Pusat Bahasa Indonesia saat ini adalah edisi 4 dan dapat diunduh secara bebas melalui situs resmi Pusat Bahasa Indonesia dalam bentuk PDF [10].

Untuk menunjang penelitian tersebut diperlukan WordNet dalam bentuk Bahasa Indonesia. Komponen dasar untuk membangun WordNet adalah konstruksi *synonym sets* karena merupakan konsep dasar untuk mendukung hubungan semantik antarkata [10]. Oleh karena itu, dalam penelitian ini dilakukan cara mengekstraksi *synonym sets* menggunakan metode *agglomerative clustering* untuk mendapatkan *synonym sets* yang optimal agar mempermudah proses pembangunan WordNet Bahasa Indonesia. *Synonym sets* yang digunakan dalam penelitian ini berasal dari *monolingual lexical resource* Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia yang dikeluarkan oleh Pusat Bahasa Indonesia.

Topik dan Batasannya

Berdasarkan latar belakang di atas, topik yang akan diangkat dalam penelitian ini adalah *clustering synonym sets* yang diambil dari Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia menggunakan metode *agglomerative clustering* untuk mengelompokkan *synonym sets* dan menggunakan metode *F-measure* untuk pengujian. Adapun batasan masalah dalam penelitian ini adalah:

- a. *Monolingual lexical resource* yang digunakan berasal dari Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia.
- b. Data yang diambil merupakan data *sample* dan sebagian kecil dari Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia.
- c. Data yang berasal dari Kamus Besar Bahasa Indonesia dan Tesaurus Bahasa Indonesia dijadikan satu dalam kandidat *synonym sets*.
- d. Metode pengujian menggunakan *F-measure* dengan membandingkan hasil *synonym sets* yang telah diproses.

Tujuan

Tujuan dari penelitian tugas akhir ini adalah mengelompokkan *synonym sets* yang memiliki kesamaan antar-kata berdasarkan *distance value* menggunakan metode *agglomerative clustering* dan pengujian hasil *clustering* menggunakan metode *F-measure* untuk mengukur akurasi hasil perbandingan *clustering* dengan validasi bersama ahli bahasa atau *lexicographer*.

Organisasi Tulisan

- a. Pendahuluan
Bab ini menjelaskan tentang latar belakang, topik dan batasan masalah, serta tujuan dari penelitian ini.
- b. Studi Terkait
Bab ini menjelaskan tentang teori dan penelitian yang telah dilakukan sebelumnya serta literatur yang mendukung dalam penelitian ini.
- c. Sistem yang Dibangun
Bab ini menjelaskan tentang proses yang dibangun pada sistem yang dihasilkan dari penelitian ini.

d. Evaluasi

Bab ini menjelaskan tentang hasil pengujian dan analisis pengujian.

e. Kesimpulan

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari penelitian ini beserta saran untuk penelitian selanjutnya.