

BAB I

Latar Belakang

WordNet adalah sebuah kamus elektronik berskala besar untuk Bahasa Inggris yang dibuat pada tahun 1986 di Universitas Princeton yang dikembangkan oleh George A. Miller, terinspirasi oleh eksperimen dalam Kecerdasan Buatan yang mencoba memahami memori semantik manusia. *WordNet* atau yang dikenal dengan Princeton *WordNet*, dikembangkan oleh seorang ahli leksikographer yang hasilnya dijadikan sebuah basis data leksikal. Princeton *WordNet* dibuat secara manual dan membutuhkan banyak sumber daya baik ahli bahasa dan waktu sehingga menghasilkan *WordNet* dengan kualitas yang tinggi. Berdasarkan penelitian lain, *WordNet* berisikan sejumlah 155.000 kata benda (nouns), kata sifat (*adjectives*), kata keterangan (*adverbs*), dan kata kerja (*verbs*), kemudian kata-kata tersebut dikelompokkan berdasarkan maknanya ke dalam *synonym sets* (synsets) atau kumpulan sinonim yang memiliki makna sama. Dalam perkembangannya *WordNet* telah dibuat dalam beberapa bahasa, diantaranya adalah *Persian WordNet* yang dibangun menggunakan sumber leksikal yang tersedia. Pada *WordNet* terdapat struktur yang berisi informasi kata, kelas kata, dan definisi dari seluruh himpunan kata yang terdapat dalam suatu bahas yang kemudian menjadi entitas tunggal yang saling berelasi. Umumnya satuan terkecil dalam kamus bahasa adalah kata, hal ini berbeda dengan *WordNet* karena satuan terkecilnya adalah synsets atau himpunan sinonim yang memiliki makna yang sama. Synsets merupakan konsep dasar yang mendukung hubungan-hubungan semantik di basis data leksikal. Pada tahun 2010 pernah dibangun sebuah synsets untuk *WordNet* Bahasa Indonesia menggunakan kamus bahasa yang dikembangkan di Institut Teknologi Sepuluh Nopember menggunakan teknik *hierarchical clustering*. Berdasarkan hasil penelitian yang dilakukan dibutuhkan teknik *clustering* yang lebih kuat dan lebih sesuai untuk pembangunan synsets untuk *WordNet* Bahasa Indonesia. Dalam penelitian ini dipilih teknik clustering *ROCK (Robust Clustering Using Links)* menggunakan ukuran kemiripan yang disebut dengan *link* dalam membentuk cluster. Teknik ini merupakan pengembangan dari teknik *hierarchical clustering* dan dirasa sangat sesuai dengan kebutuhan dalam pembangunan *synonym sets* yang mengedepankan kesamaan makna dalam pembangunannya. Berdasarkan penelitian sebelumnya penggunaan *ROCK clustering* sangat direkomendasikan dalam pembangunan synsets Bahasa Indonesia, karena teknik *clustering* ini sangat mendukung untuk melakukan proses klaterisasi dengan menggunakan atribut kategori. Kelebihan ini lah yang akan dimanfaatkan agar sistem yang dibangun mampu menghasilkan synsets yang lebih baik karena synsets merupakan komponen yang penting dalam pembangunan *WordNet*. Maka dari itu, pembangunan *synonym sets* dengan menggunakan *ROCK clustering* pada penelitian ini diharapkan mampu menghasilkan sistem yang dapat memberikan performansi atau akurasi yang lebih baik dari penelitian-penelitian sebelumnya dalam menghasilkan synsets untuk *WordNet* Bahasa Indonesia.