

ABSTRAK

Pengukuran *semantic relatedness* merupakan suatu pekerjaan untuk memperkirakan kedekatan arti dari pasangan kata. Pekerjaan ini dengan mudah bisa dilakukan oleh manusia berdasarkan pengalaman dan wawasannya, namun komputer hanya bisa melakukannya dengan mengukur nilai keterkaitannya dengan bantuan pemahaman arti kata seperti kamus, tesaurus, dan ensiklopedia. Salah satu contoh bantuannya yang sering dipakai pada pekerjaan ini yaitu WordNet yang akan digunakan pada pengerjaan tugas akhir ini. Pekerjaan ini akan mempermudah pekerjaan *Natural Language Processing*, seperti pendeteksi plagiarisme antara dua data teks yang berbeda.

Pada tugas akhir ini pengukuran *semantic relatedness* dilakukan pada sebuah model *graph* yang dibangun dengan bantuan WordNet. Setiap kata dan arti-artinya di dalam WordNet akan dijadikan sekumpulan *node* yang saling terhubung melalui beberapa tipe *edge* berdasarkan suatu relasi yang menghubungkannya. Tiap tipe *edge* tersebut memiliki cara pembobotan yang berbeda-beda. Pembobotan tersebut akan mempengaruhi hasil dari pengukuran *semantic relatedness*.

Graph yang telah dibangun tersebut adalah untuk menjalankan metode pengukuran *semantic relatedness* yang dinamakan *random walks*. *Graph* akan menjadi tempat berjalannya partikel yang berasal dari *random walks*. Partikel tersebut akan berjalan dari *node* ke *node* lainnya yang terhubung berdasarkan probabilitas berpindahannya. Probabilitas kepindahannya tersebut hanya dipengaruhi oleh *node* yang sedang disinggahi partikel, hal ini lah yang membuat *random walk* ini disebut sebagai *Markov chain*.

Hasil dari *semantic relatedness* melalui sistem yang dibuat dengan menggunakan *dataset* Rubenstein dan Goodenough mendapatkan hasil evaluasi menggunakan *Pearson correlation coefficient* dengan nilai tertinggi 0.477.

Kata kunci: *semantic relatedness, WordNet, Natural Language Processing, random walks, graph, Markov chain, Pearson correlation coefficient*