

1. Pendahuluan

1.1 Latar Belakang

Dewasa ini teknologi berkembang sangat pesat, salah satu bidangnya yaitu teknologi informasi. Berbagai cara untuk mendapatkan informasi yang ada, seperti mendapatkan informasi seberapa besar nilai kesamaan semantik antar sepasang kata dalam suatu dokumen. Kesamaan semantik memiliki peran penting dalam beberapa *task* dari *Natural Language Processing* dan beberapa bidang terkait seperti *text classification*, *document clustering*, *text summarization*, dan lain sebagainya [4]. Kesamaan semantik adalah *task* pada *Natural Language Processing* (NLP) untuk mengukur kesamaan atau kedekatan antara pasangan kata secara semantik. Pasangan kata dikatakan memiliki kesamaan semantik jika pasangan tersebut memiliki makna atau konsep yang sama. Penelitian ini di dasari dimana komputer belum dapat menyamakan persepsi manusia terkait penilaian dari makna pasangan kata yang memiliki kesamaan semantik. Atas dasar ini lah,tugas akhir dibentuk agar komputer dapat menghasilkan informasi seberapa besar nilai kesamaan semantik antar sepasang kata dalam suatu dokumen. Untuk mencari atau menghitung kesamaan semantik antar kata ada beberapa metode yang bisa digunakan salah satunya menggunakan metode berbasis vektor. GloVe merupakan suatu metode *unsupervised learning* pada representasi kata yang mengungguli model lainnya di *word analogy*, *word similarity*, dan *named entity recognition* [8]. Sebagian besar metode vektor bergantung pada jarak atau sudut di antara vektor pasangan kata sebagai metode utama untuk mengevaluasi kualitas representasi kata tersebut. Penggunaan metode berbasis vektor memberikan kinerja mutakhir untuk kesamaan semantik dan metode berbasis vektor dengan mudah diadaptasi di tiga basis pengetahuan yang berbeda (WordNet, Wikipedia, dan Wiktionary) [11]. Hal ini menunjukkan bahwa metode berbasis vektor dapat menjadi pilihan yang baik dan metode Glove umumnya menghasilkan nilai kesamaan yang tinggi. Dalam metode berbasis vektor ini untuk mengukur kesamaan semantik antar sepasang kata menggunakan fungsi kesamaan kosinus. Fungsi kesamaan kosinus adalah fungsi yang umumnya digunakan pada sebagian besar metode berbasis vektor untuk menghitung nilai kesamaan antar dua vektor kata. Pada penelitian tugas akhir ini diimplementasikan metode GloVe untuk mengukur kesamaan antar pasangan kata menggunakan korpus Wikipedia bahasa Indonesia dan skor yang dihasilkan akan dihitung korelasinya menggunakan *correlation pearson* dengan membandingkan skor *gold standard* dari WordSim353, Sim-Lex999 dan Miller Charles.

1.2 Topik dan Batasannya

Pada penelitian ini memiliki rumusan masalah sebagai berikut :

1. Bagaimana mengimplementasikan kesamaan semantik sepasang kata dalam bahasa Indonesia dengan menggunakan metode GloVe?
2. Bagaimana hasil Analisis nilai korelasi yang dihasilkan menggunakan metode GloVe dengan perbandingan dataset *gold standard* WordSim-353,SimLex-999, dan Miller Charles?

Adapun batasan dari penelitian ini adalah korpus Wikipedia bahasa Indonesia sebagai dataset yang digunakan dalam pengukuran kesamaan semantik. Korpus Wikipedia ini berasal dari hasil *crawling* bapak Herry Sujaini, beliau adalah seorang dosen Informatika pada Universitas TanjungPura dan dataset yang digunakan sebagai perbandingan untuk melihat sistem yang dibangun sudah baik atau belum menggunakan dataset *gold standard* WordSim-353, SimLex-999, dan Miller Charles. Dataset *Gold standard* ini berasal dari penelitian sebelumnya menggunakan metode PMI [7].

1.3 Tujuan

Berdasarkan rumusan masalah yang telah disampaikan, berikut tujuan penelitian ini adalah

1. Menerapkan metode GloVe untuk mengimplementasikan kesamaan semantik antar sepasang kata dalam bahasa Indonesia.
2. Menganalisis nilai korelasi kesamaan semantik pasangan kata dalam bahasa Indonesia menggunakan metode GloVe dengan perbandingan dataset *gold standard* WordSim-353, SimLex-999, dan Miller Charles.

1.4 Organisasi Tulisan

Organisasi tulisan pada penelitian ini terdiri atas : Bagian 1 membahas mengenai latar belakang, rumusan masalah dan tujuan dari penelitian. Bagian 2 membahas studi terkait mengenai teori. Bagian 3 membahas sistem yang dibangun untuk metode GloVe. Bagian 4 menjelaskan hasil dan analisis sistem. Bagian 5 menjelaskan kesimpulan serta saran.