

BAB VI

PENUTUP

6.1 Kesimpulan

Dari hasil perancangan dan implementasi sistem maka dapat disimpulkan:

1. Algoritma KNN (*K-Nearest Neighbor*) berbasis ontologi dapat diterapkan pada proses pengklasifikasian berita berbahasa Inggris. Sebelum dilakukan proses KNN dilakukan proses *preprocessing* dimana di dalam proses *preprocessing* ini terdapat proses *ontology extraction* dan *feature selection* untuk mereduksi vektor *term*. Proses KNN dilakukan dengan memberikan skor pada setiap kategori setelah sebelumnya dokumen uji melakukan *learning* dengan setiap dokumen latih dan menghasilkan nilai *cosine similarity* antara dokumen latih dengan dokumen uji.
2. Hasil evaluasi sistem menunjukkan rata-rata *precision*, *recall*, dan *f1-measure* tertinggi berturut-turut adalah 0.958, 0.95, dan 0.949 dengan nilai *threshold* = 1 dan nilai “k” = 12, 14, 15, 17, 18, 19 dan nilai evaluasi pada nilai “k” dan *threshold* tersebut bernilai sama. Pada setiap pengujian tersebut hasil *f1-measure* selalu diatas 0.85 dengan parameter *threshold* dan jumlah “k” berdasarkan pada pengujian sebelumnya. Data yang digunakan juga berpengaruh terhadap hasil penelitian. Pada penelitian ini beberapa data latih memiliki kemiripan antar kategori nya dilihat dari isi nya. Jika nilai “k” yang diberikan kecil maka akan mempengaruhi hasil dari klasifikasi. Berdasarkan hasil pengujian yang telah dilakukan dapat disimpulkan bahwa algoritma KNN (*K-Nearest Neighbor*) berbasis ontologi dapat melakukan klasifikasi dokumen dengan baik.

6.2 Saran

Sistem yang telah dibangun pada penelitian ini masih memiliki kekurangan. Adapun beberapa saran untuk penelitian lebih lanjut adalah sebagai berikut:

1. Beberapa *term* tidak dapat digabungkan di karenakan *synset* yang digunakan hanya terbatas pada *synset noun* saja. Proses *preprocessing* khususnya proses *ontology extraction* dapat dikembangkan dengan menambahkan *synset verb*, *adjective*, dan *adverb* sehingga memperkaya daftar *synset* yang digunakan untuk *ontology extraction*.
2. Data teks yang digunakan merupakan berita umum, dimana kata-kata yang ada didalamnya juga merupakan kata-kata umum. *Database WordNet* tidak dapat menemukan hubungan semantik dari kata-kata pada bidang tertentu. Jika data yang digunakan merupakan teks pada bidang tertentu seperti jurnal ilmiah dengan topik kedokteran atau bidang ilmu lainnya, dimana terdapat kata-kata yang spesifik mengenai bidang ilmu tersebut maka kita dapat merancang sendiri basis ontologi untuk menangani data tersebut.

