

Abstraksi

Kategorisasi teks merupakan sebuah otomatisasi dalam pengkategorian teks berdasarkan konten yang terdapat dalam dokumen tersebut. Pada riset kategorisasi teks sebelumnya, dokumen direpresentasikan sebagai “*bag of words*” yaitu sekumpulan kemunculan *term* dalam dokumen tersebut. Dalam melakukan kategorisasi teks, pertimbangan frekuensi kemunculan *term* saja tidak cukup sebab dalam melakukan proses kategorisasi teks perlu pula mempertimbangkan susunan *term* dalam dokumen. Oleh sebab itu digunakan skema pembobotan *Distributional Feature*. Pertimbangan *Distributional Feature* yang dibahas dalam tugas akhir ini adalah *compactness of appearance* dan *position of the first appearance of a word*.

Dalam Tugas Akhir *Distributional Feature* digabungkan dengan skema pembobotan *Term Frequency Inverse Document Frequency* (TFIDF) dan *Inverse Term Frequency* (ITF). Kemudian dilakukan perbandingan performansi berupa nilai *precision*, *recall*, dan *F1 measure* antara TFIDF, TFIDF·DF (TFIDF dan *Distributional Feature*) dan ITF·DF (ITF dan *Distributional Feature*).

Hasil pengujian menunjukkan bahwa, penggunaan *Distributional Feature* dengan skema pembobotan TFIDF (TFIDF·DF) dan *Distributional Feature* dengan ITF (ITF·DF) mampu menghasilkan performansi yang cukup baik. Hal ini dibuktikan dengan rata-rata *precision*, *recall*, dan *F1 measure* yang mendekati angka 1. Selain itu, penggunaan *Distributional Feature* dengan skema pembobotan TFIDF dan ITF mampu menghasilkan performansi yang lebih baik jika dibandingkan dengan skema pembobotan TFIDF ketika menggunakan 25%, 50%, dan 75% dataset dari setiap kelas untuk masing-masing topik. Sedangkan pada saat menggunakan 100% dataset, TFIDF menghasilkan performansi yang lebih baik dibandingkan dengan skema pembobotan TFIDF·DF dan ITF·DF.

Kata kunci: kategorisasi teks, *Distributional Feature*, TFIDF, TFIDF·DF, ITF·DF, performansi