

Abstrak

Tindak plagiarisme bisa saja dilakukan tanpa sengaja. Kemungkinan itu bisa terjadi karena kurangnya informasi seseorang terhadap karya tulis orang lain yang memang sudah dipublikasi terlebih dahulu. Oleh karena itu, kreator bertanggung jawab untuk memeriksa apakah karyanya terindikasi plagiat atau tidak sebelum dipublikasi. Masalah muncul, tidak memungkinkan untuk seseorang memeriksa satu persatu ke semua karya tulis yang ada. Karenanya dibutuhkan sistem yang mampu mendeteksi indikasi plagiat secara otomatis. Pada penelitian ini, sistem dibangun dengan menggunakan pendekatan *machine learning*, yaitu dengan menggunakan *Bayesian network*. Untuk mencapai pendekatan ini, dibutuhkan ekstraksi *feature* berupa *unigram*, *POS gram*, dan *N-gram*. *Feature-feature* tersebut menangani kasus leksikal dan sintaksis, yang mana memiliki peranan penting sebagai penanda indikasi plagiarisme. Untuk mengukur performansi sistem yang dibangun, digunakan *F1-score*. Sistem yang dibangun mampu mendeteksi indikasi plagiat pada dokumen berbahasa Inggris, dengan *F1-score* tertinggi mencapai 0.987.

Kata kunci : *Plagiarisme; Bayesian Network; Unigram; POS Gram; N-gram.*