

Abstrak

Dalam membuat suatu karya tulis atau karya ilmiah tidak dapat dipungkiri bahwa penulis membutuhkan informasi dan referensi dari hasil karya orang lain. Referensi tersebut bisa didapatkan melalui sebuah artikel, buku, paper, atau hasil karya lainnya yang memiliki keterkaitan dengan referensi yang dibutuhkan. Seiring berjalannya waktu, era digital semakin berkembang dan mengakibatkan suatu dokumen mulai direpresentasikan dalam bentuk digital dan dapat di sebarluaskan dengan mudah dengan adanya internet. Hal tersebut mengakibatkan tindakan plagiarisme cenderung lebih mudah dilakukan. Dalam paper ini akan dilakukan pendeteksian plagiarisme menggunakan pendekatan *text alignment* dengan menggunakan proses *Preprocessing, Seeding, Extension* dan *Filtering*. Tujuan dari *text alignment* merupakan mencari pasangan fragmen yang digunakan kembali atau menjadi sumber dari dokumen *suspicious*. Dataset yang digunakan didapatkan dari PAN dengan *task plagiarism detection* yang terkategori menjadi beberapa tipe yaitu *no plagiarism, no obfuscation, random obfuscation, translation obfuscation, dan summary obfuscation* dimana setiap tipe memiliki karakteristiknya masing masing. Adapun dalam penyelesaiannya akan digunakan metode *sentence similarity* menggunakan TF-IDF, *cosine similarity, dice coefficient* dan kesamaan semantik yang kemudian hasil akhir dari fragmen yang dihasilkan akan dievaluasi dengan dua evaluasi yaitu evaluasi level kasus dan evaluasi level karakter. Untuk evaluasi pada level kasus didapatkan nilai *f-measure* sebesar 0.9768 sedangkan untuk evaluasi level karakter didapatkan nilai *f-measure* sebesar 0.7090. Apabila diurutkan terhadap hasil dari penelitian lainnya berdasarkan nilai *recall* yang dihasilkan mendapatkan peringkat 11 dari 16 partisipan, sedangkan apabila diurutkan berdasarkan nilai *precision* mendapatkan peringkat 17 dari 19 partisipan.

Kata kunci : plagiarisme, *text alignment, sentence similarity, cosine similarity, dice similarity, TF-IDF, parameter adaptif.*