ABSTRAK

Dokumentasi laporan skripsi dilakukan dengan menyimpan seluruh berkas digital dalam basis data yang disebut dengan repository eprint. Mekanisme pencarian eprint di repository jurusan informatika UPN Veteran Yogyakarta masih dilakukan dengan cara SQL. Teknik pencarian dengan model SQL membatasi pencarian secara *lexical* sehingga hasilnya terpaku pada kata kunci yang diberikan. Hal ini menyulitkan pengguna dalam mencari eprint yang relevan. Pendekatan text similarity dapat membantu proses pencarian.

Algoritma text similarity yang akan dibahas di penelitian ini yaitu TF-IDF dan Fasttext. Fitur yang dihasilkan dari TF-IDF dan Fasttext kemudian digunakan untuk kalkulasi kedekatan dengan model perhitungan cosine similarity. Kalkulasi dari cosine similarity dijadikan dasar rangking pencarian yang kemudian dibandingkan hasilnya antara kedua algoritma TF-IDF dan Fasttext.

Pengujian dengan metrik pengukuran *mean reciprocal rank* (MRR), akurasi algoritma TF-IDF memperoleh 94.92%, 89.16% dan 80.69% sedangkan algoritma Fasttext memperoleh 93.84%, 85.65% dan 79.40%. Pengujian lain dengan metrik pengukuran *root mean square error* (RMSE) didapatkan algoritma TF-IDF sebesar 27.37%, 44.52% dan 54.79% sedangkan algoritma Fasttext memperoleh nilai RMSE 6.34%, 11.80% dan 15.41%. Hasil ini menunjukan bahwa TF-IDF memiliki tingkat akurasi yang sedikit lebih baik dibandingkan Fasttext meskipun nilai *cosine similarity* TF-IDF relatif lebih rendah daripada Fasttext.

Kata Kunci : Text Similarity, Information Retrival, TF-IDF, Fasttext