Abstrak

Sistem information retrieval adalah suatu sistem yang terdiri dari beberapa tahapan dalam usaha menemukan dokumen yang relevan dalam kumpulan dokumen. Pada umumnya *sistem information retrieval* menggunakan stemmer untuk mereduksi berbagai variasi morfologi yang terdapat pada dokumen sumber. Karena terkait dengan bahasa dokumen yang digunakan, maka umumnya *stemmer* yang digunakan dibuat berdasarkan aturan bahasa tertentu. Sedangkan bahasa dokumen yang digunakan bisa saja lebih dari satu bahasa.

Sistem ini mengimplementasikan n-gram *stemmer* yang merupakan sebuah metode *stemming* yang tidak terikat dengan satu aturan bahasa apapun. Sebagai bahan uji, bahasa yang digunakan dalam dokumen terdiri dari bahasa Inggris dan Prancis. Selain itu sebagai pembanding dengan *stemmer* yang umum digunakan pada satu bahasa digunakan porter stemmer.

Hasil pengujian yang telah dilakukan pada Tugas Akhir ini menunjukkan penerapan n-gram *stemmer* bisa dilakukan terhadap dokumen lebih dari satu bahasa. Kekuatan kompresi dan akurasi yang dihasilkan n-gram *stemmer* jika dibanding dengan porter maka hasilnya bervariasi tergantung nilai n pada n-gram yang digunakan.

Kata kunci: sistem *information retrieval*, *stemming*, n-gram *stemmer*