

Abstrak

Berkembangnya teknologi di dunia maya membuat jumlah informasi berupa artikel semakin banyak. Untuk itu, diperlukan suatu metode terhadap artikel yang memudahkan pembaca mencari informasi dengan menerapkan salah satu fungsionalitas dari data mining, yaitu kategorisasi. Untuk mendapatkan hasil kategorisasi yang baik diperlukan tahap *preprocessing data* yang baik pula. Salah satu tahap *preprocessing data* yang umum digunakan adalah *stemming*.

Stemming adalah proses untuk menemukan akar kata atau kata dasar dengan memisahkan semua *affix* atau imbuhan yang melekat pada kata tersebut. Dengan proses *stemming* ini akan mampu mengurangi dimensi dari data dalam melakukan proses kategorisasi sehingga dapat meningkatkan hasil kategorisasi karena beberapa kata yang memiliki kata dasar yang sama dapat dikelompokkan menjadi satu *stem*. Ada beberapa metode *stemming* menurut prosesnya, salah satunya adalah *affix removal*. Pada Tugas Akhir ini akan dibahas beberapa metode *affix removal* yaitu Porter Stemmer dan Krovetz Stemmer, serta pengaruhnya terhadap proses kategorisasi. Performansi *stemmer* dihitung berdasarkan nilai akurasi (*accuracy*) dan ICF (*Index Compression Factor*).

Berdasar hasil pengujian, Modifikasi Porter Stemmer memiliki nilai performansi yang lebih baik dibanding Porter Stemmer dan Krovetz Stemmer. Namun dari hasil performansi *stemmer* terbaik yang didapat oleh Improved Porter Stemmer belum tentu bisa meningkatkan nilai *precision* dan *recall* dalam kategorisasi.

Kata kunci: *stemming*, *affix removal*, Porter Stemmer, Krovetz Stemmer, *Index Compression Factor*, *accuracy*.