ABSTRAK

PENCARIAN KATA TIDAK BAKU PADA DOKUMEN DI BALAI BAHASA BANDUNG MENGGUNAKAN *STEMMING ENGINE* BAHASA INDONESIA

Oleh

JAJANG APRIANSYAH

10108188

Balai Bahasa Bandung adalah lembaga yang bergerak dibidang kebahasaan dan kesastraan dan berada di bawah naungan Pusat Bahasa. Salah satu program kerja Balai Bahasa Bandung yaitu mengumpulkan karya ilmiah atau karya tulis dan mencetaknya dalam satu buku, untuk melakukan pemeriksaan karya ilmiah atau karya tulis dan pendistribusian karya penulis maka membutuhkan aplikasi yang bisa membantu proses pemeriksaan dokumen yang mengandung kata tidak baku. Hal ini dapat diatasi dengan *stemming engine*.

Stemming adalah proses pemisahan kata dengan imbuhannya sehingga menghasilkan sebuah kata dasar. Ada dua proses yang harus dilakukan sebelum *stemming* yaitu *tokenizing* dan *filtering*. *Tokenizing* adalah proses pemotongan kalimat menjadi per kata, sedangkan *filtering* adalah proses penghilangan kata yang dianggap tidak penting untuk dilakukan *stemming*.

Berdasarkan hasil penelitian dan setelah dilakukan pengujian terhadap *stemming engine*, kesimpulan yang dapat diambil yaitu *stemming engine* dapat membantu proses pemeriksaan dokumen yang mengandung kata tidak baku di Balai Bahasa Bandung, dapat membantu mempercepat proses pemeriksaan dan dapat membantu pendistribusian karya ilmiah ataupun karya tulis ke Balai Bahasa Bandung.

Kata Kunci: *stemming*, *tokenizing*, *filtering*, bahasa, Indonesia