

Abstrak

Part-of-speech tagging merupakan proses pemberian kelas kata terhadap suatu kata dalam sebuah teks. Sebuah POS Tagger untuk bahasa tertentu umumnya dibangun dengan menggunakan korpus dalam domain umum, misalnya korpus yang didapat dari teks surat kabar. Apabila POS Tagger ini digunakan pada kata dari domain yang baru atau lebih spesifik, maka POS Tagger tersebut dapat memberikan kelas kata yang tidak akurat. Untuk menyelesaikan permasalahan adaptasi domain spesifik dapat dilakukan beberapa metode, antara lain menggunakan *clustering* untuk mengubah representasi kata atau menggunakan model dengan leksikon berjumlah besar dan menggunakan teks yang berlabel dari domain spesifik untuk pelatihan model. Pada tugas akhir ini diterapkan metode adaptasi domain dengan menggunakan leksikon tambahan yang dibangun berdasarkan aturan imbuhan. Domain spesifik yang diadaptasi adalah domain produk kecantikan. Komponen yang dibutuhkan adalah sebuah POS Tagger Indonesia dengan domain umum dan leksikon dari domain target yang belum dilabeli. Pemberian kelas kata pada leksikon domain target diberikan berdasarkan informasi afiks dan sisanya dilabeli secara manual. Berdasarkan observasi dataset, terdapat banyak penggunaan bahasa Inggris sehingga leksikon dibangun dalam bahasa Indonesia dan bahasa Inggris. Leksikon yang sudah diproses ditambahkan ke dalam leksikon dari POS Tagger untuk memberikan informasi domain spesifik kepada POS Tagger dengan domain umum. POS Tag yang difokuskan pada tugas akhir ini adalah kelas kata benda, kata sifat dan kata keterangan karena hasil dari POS Tagger digunakan untuk melakukan ekstraksi aspek dan opini. Tagger yang diberikan leksikon tambahan memperoleh akurasi sebesar 68.55%. Selain itu, persentase kata yang memiliki label atau tidak berlabel X menjadi 92.17%.

Kata kunci : POS Tagger, adaptasi domain, leksikon, informasi afiks