

Pengembangan part of speech tagger untuk bahasa Indonesia berdasarkan metode conditional random fields dan transformation based learning

Triastuti Chandrawati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123614&lokasi=lokal>

Abstrak

Part of Speech Tagging (POS Tagging) adalah kegiatan pemberian label kelas kata pada suatu kata. Proses ini pada awalnya dilakukan secara manual. Namun proses part of speech tagging secara manual menghabiskan banyak waktu dan tenaga karena membutuhkan banyak ahli bahasa untuk memberikan tag pada setiap kata. Masalah ini kemudian menjadi pendorong bagi para peneliti untuk membangun metode dan aplikasi yang dapat melakukan part of speech tagging secara otomatis. Penelitian di bidang part of speech tagger untuk Bahasa Indonesia belum terlalu mendapat perhatian, karena sampai saat ini, belum ada suatu aplikasi part of speech tagger untuk Bahasa Indonesia. Maka penulis melakukan penelitian untuk menghasilkan suatu aplikasi part of speech tagger untuk Bahasa Indonesia dengan memanfaatkan tiga metode yaitu Conditional Random Fields, Transformation Based Learning, dan kombinasi kedua metode ini. Penelitian ini menggunakan korpus Bahasa Indonesia yang tersusun atas 49 buah artikel surat kabar dan terdiri atas total 13.465 buah token. Tagset yang digunakan pada penelitian ini terdiri atas 21 jenis tag (21 jenis kelas kata). Tagset ini merupakan modifikasi dari Penn Treebank Tagset.