

Perbandingan decision tree, maximum entropy, dan association rules pada resolusi koreferensi untuk bahasa Indonesia

Astria Kurniawan Sumantri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123215&lokasi=lokal>

Abstrak

Resolusi koreferensi merupakan suatu proses yang dilakukan untuk mengetahui hubungan antara suatu entitas bernama atau kata ganti dengan entitas bernama lainnya dalam suatu teks. Hubungan tersebut dapat diketahui berdasarkan informasi kelas kata, kelas nama, kesamaan karakter penyusun kata, dan letak entitas bernama atau kata ganti tersebut dalam suatu teks. Penelitian ini membandingkan penerapan resolusi koreferensi menggunakan pendekatan machine learning dengan metode decision tree, maximum entropy, dan association rules. Penelitian dilakukan untuk mengetahui metode mana yang lebih baik untuk diaplikasikan pada tugas resolusi koreferensi untuk Bahasa Indonesia. Selain itu penelitian ini dilaksanakan untuk mengetahui efektifitas penggunaan fitur pada resolusi koreferensi untuk Bahasa Indonesia. Tujuan lainnya adalah mengetahui berapakah jumlah dokumen pelatihan yang diperlukan untuk masing-masing metode sehingga sistem memberikan hasil yang optimal. Uji coba dilakukan menggunakan 500 dokumen yang dikumpulkan dari Kompas Online (www.kompas.com). Secara umum, nilai recall metode decision tree lebih tinggi dari metode yang lain. Hasil uji coba memperlihatkan bahwa association rules menghasilkan kinerja yang paling baik dengan F-measure sebesar 81,36% yang diamati pada jumlah dokumen pelatihan 300 dan menggunakan semua fitur. Metode decision tree mencapai kestabilan pada jumlah dokumen 50. Metode maximum entropy mencapai kestabilan pada jumlah dokumen 70, sedangkan metode association rules mencapai kestabilan saat jumlah dokumen 90. Pada metode decision tree dan maximum entropy, kombinasi fitur minimum yang memberikan hasil terbaik adalah kombinasi fitur yang berhubungan dengan kesamaan karakter penyusun frase. Sedangkan untuk metode association rules, kombinasi fitur yang berhubungan dengan kesamaan karakter penyusun frase dan kombinasi fitur kelas nama merupakan kombinasi yang memberikan hasil terbaik.