

Pengenalan entitas bernama menggunakan metode association rules pada dokumen berbahasa Indonesia

Markus, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=124165&lokasi=lokal>

Abstrak

Pengenalan entitas bernama merupakan tugas dasar dalam sistem ekstraksi informasi untuk mengenali entitas nama, entitas waktu, dan entitas bilangan. Entitas nama meliputi nama organisasi, nama orang, dan nama lokasi. Pengenalan entitas bernama dilakukan dengan menggunakan aturan yang dibentuk dari sekumpulan fitur suatu kata pada metode association rules. Pencocokan fitur-fitur kata dengan aturan dapat dilakukan dengan dua cara, exact match dan partial match. Pada partial match, digunakan nilai similarity yang merepresentasikan kesamaan suatu fitur-fitur kata dengan aturan. Fitur-fitur yang digunakan dalam penelitian ini dibagi menjadi empat bagian. Fitur pertama merupakan informasi orthographical suatu token (term feature), fitur kedua berisi informasi orthographical token sebelum dan sesudah (neighbourhood feature), fitur ketiga berisi informasi token dalam suatu daftar (list feature), dan fitur keempat berisi informasi global suatu token (global feature). Dokumen yang digunakan pada penelitian ini berjumlah 160 dokumen dengan perbandingan 100 dokumen untuk pelatihan dan 60 dokumen untuk pengujian. Dokumen-dokumen tersebut diperoleh dari koran online berbahasa Indonesia, yaitu Kompas dan Republika. Uji coba dilakukan pada beberapa kombinasi penggunaan keempat fitur di atas, variasi nilai similarity dari 0,1 sampai 0,9 pada partial match, dan variasi jumlah dokumen pelatihan dari 10 dokumen sampai 100 dokumen dengan penambahan jumlah dokumen sebanyak 10 dokumen. Jumlah dokumen pengujian yang digunakan berjumlah 60 dokumen. Hasil uji coba menunjukkan bahwa sistem dapat mengenali entitas bernama dengan nilai F-measure tertinggi 43,34% dengan menggunakan term feature, neighbourhood feature, dan list feature dengan nilai similarity 0,7 pada metode partial match dengan menggunakan 100 dokumen pelatihan.