

Ekstraksi Sinonim Set dengan Agglomerative Hierarchical Clustering pada Wordnet Bahasa Indonesia

Gilang Rananda Tama, Mochammad Arif Bijaksana

Fakultas Informatika, Universitas Telkom, Bandung

gilangrananda@students.telkomuniversity.ac.id, arifbijaksana@telkomuniversity.ac.id

Abstrak

Wordnet Indonesia adalah kamus leksikal yang dikembangkan oleh Retrieval Information Lab, Fakultas Ilmu Komputer, Universitas Indonesia. Saat ini wordnet bahasa Indonesia bisa dikatakan jauh dari kata sempurna dan masih membutuhkan pengembangan serta kosakata yang terdapat didalamnya relatif sedikit. Karena perkembangan Wordnet Bahasa Indonesia ini dianggap tidak terlalu menarik menyebabkan jumlah penelitian tentang pemrosesan bahasa alami masih sangat minim. Pengembangan yang akan dilakukan penulis disini dengan mengimplementasikan metode Agglomerative Hierarchical Clustering sebagai sarana untuk memproses data serta mengetahui performansi dalam pembangunan Wordnet Bahasa Indonesia. Secara sederhana cara kerja dari metode ini dengan mengelompokkan kata yang mempunyai kemiripan dalam arti dan makna kemudian dimasukkan ke dalam satu klaster, setelah itu dilakukan pengecekan lagi apakah ada kata lain yang memiliki makna yang sama, apabila tidak maka itu akan menjadi satu klaster yang didalamnya terdapat kata dengan makna yang sama atau bisa juga disebut synonym set. Untuk dataset, penulis mengambil data dari Tesaurus Standar Bahasa Indonesia. Data hasil penelitian ini diperoleh secara manual dan acak. Untuk setiap stepnya pertama hitung matrik jarak antar data guna mengetahui kata mana yang memenuhi syarat untuk dimasukkan ke dalam klaster, kemudian gabungkan 2 kelompok terdekat berdasarkan parameter kedekatan yang sudah dicari sebelumnya kemudian perbarui matrik kedekatan tadi untuk mempresentasikan kedekatan kelompok baru dengan kelompok yang masih tersisa. Lakukan pengulangan secara terus menerus hingga hanya 1 kelompok yang tersisa. Dan nantinya mendapatkan hasil berupa beberapa klaster yang bisa menjadi acuan untuk pengembangan Wordnet Bahasa Indonesia.

Kata kunci : WordNet, Agglomerative Hierarchical Clustering, Tesaurus Bahasa Indonesia
