

Pengembangan versi awal wordnet bahasa Indonesia dengan pendekatan expand menggunakan aplikasi web-based dan analisis statistik KAPPA

Desmond Darma Putra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123047&lokasi=lokal>

Abstrak

WordNet merupakan basis data kamus bahasa Inggris yang dikembangkan oleh Princeton University. WordNet memfokuskan isinya lebih kepada makna kata daripada kata itu sendiri. Saat ini, WordNet banyak digunakan sebagai sumber informasi untuk aplikasi di bidang Information Retrieval dan Natural Language Processing. Selain itu, WordNet juga mulai dikembangkan untuk bahasa-bahasa lain seperti Jerman, Perancis, Belanda, dan lain lain. Secara umum, pengembangan WordNet dapat dibagi menjadi dua, yaitu pendekatan merge dan expand.

Dalam penelitian tugas akhir ini, penulis mengembangkan versi awal untuk WordNet bahasa Indonesia dengan menggunakan pendekatan expand. Data-data yang dipersiapkan untuk pengembangan WordNet bahasa Indonesia, antara lain Princeton WordNet 3.0, kamus dwibahasa (Inggris-Indonesia) dan Kamus Besar Bahasa Indonesia (KBBI). Langkah pertama yang dilakukan dengan pendekatan ini adalah memetakan makna yang terdapat di dalam Princeton WordNet dengan makna yang terdapat dalam KBBI. Makna yang digunakan dari Princeton WordNet untuk pemetaan ini difokuskan kepada makna yang merupakan anggota Common Base Concepts. Proses ini dilakukan dengan menggunakan aplikasi web-based dan melibatkan sejumlah pengguna untuk berpartisipasi. Langkah selanjutnya yang dilakukan adalah mengevaluasi hasil pemetaan tersebut dengan memeriksa kelas kata dan kesepakatan antar pengguna dengan menggunakan statistik Kappa. Setelah proses evaluasi, dibuatlah basis data WordNet bahasa Indonesia. Basis data ini dibuat dengan mengadaptasi struktur basis data dan relasi semantik yang ada di dalam Princeton WordNet. Langkah terakhir yang dilakukan adalah membuat aplikasi web-based untuk mengakses WordNet bahasa Indonesia. Dari 3920 makna Princeton WordNet yang dipersiapkan, 1544 makna diantaranya berhasil dipetakan ke dalam bahasa Indonesia. Jumlah makna yang dipetakan ini dipengaruhi oleh beberapa hal, salah satunya adalah kurangnya promosi mengenai eksperimen pemetaan ini sehingga jumlah pengguna yang berpartisipasi masih sedikit.

Hasil pemetaan tersebut kemudian dievaluasi dengan memeriksa kelas kata dan kesepakatan antar pengguna (statistik Kappa). Secara rata-rata, kesepakatan antar pengguna dalam melakukan pemetaan dapat dikategorikan intermediate to good. Setelah evaluasi tersebut, dibuat 4 buah versi basis data antara lain versi tolerant, strict, very strict, dan ?strict + filter kesepakatan?. Versi "strict + filter kesepakatan" inilah yang dipilih menjadi basis data dari WordNet bahasa Indonesia. Basis data tersebut dapat diakses melalui aplikasi web-based dengan alamat <http://bahasa.cs.ui.ac.id/iwn/>.