

Penemuan jawaban pada sistem tanya jawab bahasa Indonesia-Inggris dengan pembobotan kata dan informasi dari internet

Deskripsi Dokumen: <http://lib.ui.ac.id/bo/uibo/detail.jsp?id=123207&lokasi=lokal>

Abstrak

Perolehan informasi adalah salah satu bidang di ilmu komputer yang bertujuan untuk membantu menemukan informasi tertentu di antara banyak informasi yang tersedia. Sistem tanya jawab adalah satu topik pada bidang perolehan informasi, di mana informasi yang diperlukan direpresentasikan dalam suatu kueri berbentuk kalimat tanya dan menghasilkan jawaban singkat.

Penelitian ini menghasilkan sistem tanya jawab bilingual di mana pertanyaan diberikan dalam Bahasa Indonesia dan jawaban dihasilkan dalam Bahasa Inggris.

Kueri yang digunakan adalah 190 pertanyaan Bahasa Indonesia dari topik tanya jawab Cross Language Information Retrieval (CLEF) 2006 yang terdiri dari pertanyaan fakta dan definisi. Koleksi dokumen yang digunakan adalah koleksi koran berbahasa Inggris: Glasgow Herald dan Los Angeles Times.

Untuk menemukan jawaban dari pertanyaan, pertama-tama sistem tanya jawab melakukan analisis terhadap teks pertanyaan untuk menentukan jenis jawaban yang diperlukan. Kemudian pertanyaan diterjemahkan ke Bahasa Inggris menggunakan ToggleText. Setelah itu pertanyaan yang telah berbahasa Inggris dijadikan kueri ke Lemur Toolkit untuk mendapatkan sekumpulan cuplikan dokumen dari koleksi dokumen yang relevan terhadap pertanyaan.

Bila pertanyaan bertipe fakta, maka semua entitas bernama dikumpulkan dari cuplikan-cuplikan dokumen tersebut untuk menjadi kandidat jawaban. Terdapat empat penilaian untuk memilih satu jawaban di antara para kandidat tersebut: frekuensi kemunculan kandidat jawaban yang diperoleh dari search engine; urutan dokumen yang diperoleh dari sistem perolehan informasi; jarak antarkata (Average Distance Weight); dan bobot kata kandidat jawaban.

Bila pertanyaan bertipe definisi maka perolehan jawaban dilakukan dengan mengekstrak frasa kata benda dari tiap cuplikan dokumen yang berisi kata-kata pertanyaan.

Hasil evaluasi menunjukkan bahwa sistem tanya jawab ini berhasil mencapai ketepatan jawaban sebesar 14%.