

Evaluasi teknik pengindeksan untuk perolehan informasi pembicaraan

Sutanto Sugii Joji, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123660&lokasi=lokal>

Abstrak

Semakin banyaknya dokumen pembicaraan menimbulkan kebutuhan untuk melakukan pencarian terhadap dokumen pembicaraan. Akan tetapi, metode perolehan informasi untuk dokumen teks tidak dapat langsung diterapkan pada dokumen pembicaraan. Isi dari dokumen pembicaraan adalah sinyal suara. Sinyal suara ini harus diproses terlebih dahulu agar didapat isi dari pembicaraan. Proses ini dinamakan pengenalan pembicaraan. Sistem perolehan informasi pembicaraan adalah sebuah sistem yang mengimplementasi teknik-teknik perolehan informasi dan menerapkannya kepada dokumen pembicaraan. Sinyal suara yang ada pada dokumen pembicaraan diproses terlebih dahulu dengan proses pengenalan pembicaraan agar didapat teks transkripsi pembicaraan. Sistem perolehan informasi pembicaraan melakukan pembuatan indeks berdasarkan teks pembicaraan hasil pengenalan pembicaraan.

Penelitian ini menggunakan tiga koleksi dokumen. Koleksi pertama berisi 250 dokumen pembicaraan yang didapat dari pembicaraan telepon. Koleksi ini telah dibersihkan dari noise. Koleksi kedua berisi 100 dokumen pembicaraan yang didapat dari pembicaraan radio Pro3 Radio Republik Indonesia. Koleksi ketiga berisi 29.575 dokumen yang berasal dari pembicaraan telepon. Koleksi ketiga tidak dibersihkan dari noise. Word Error Rate dari tiap-tiap koleksi adalah 26.50%, 28.40%, dan 74.20%. Teknik-teknik yang diujicobakan adalah pembuatan indeks dengan masukan hasil pengenalan pembicaraan (Transkripsi), lima alternatif kata hasil pengenalan (Transkripsi-5), alternatif kata hasil pengenalan dengan probabilitas tinggi (Transkripsi-AB), pemetaan kata hasil pengenalan terhadap kamus pengucapan untuk mendapatkan rangkaian fonem (Transkripsi-Fonem), dan rangkaian fonem 3-gram (Transkripsi-3-gram). Selain dengan pengenalan pembicaraan, penelitian ini juga mencoba memberikan variasi pembuatan indeks dengan masukan dari hasil pengenalan pembicaraan yang telah dimodifikasi agar dapat mengenali fonem (Fonem). Setelah mendapatkan hasil fonem, rangkaian 3-gram juga dibuat dari hasil fonem yang didapat (Fonem-3-gram). Penelitian ini juga melakukan eksperimen dengan cara penggabungan indeks kata dan indeks.