

Sistem question answering bahasa indonesia untuk pertanyaan non-factoid./ Novi Yusliani, Ayu Purwarianti

Novi Yusliani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20448586&lokasi=lokal>

Abstrak

Fokus dari penelitian ini adalah untuk mengembangkan data dan sistem Question Answering (QA) Bahasa Indonesia untuk pertanyaan non-factoid. Penelitian ini merupakan penelitian QA non-factoid pertama untuk Bahasa Indonesia. Adapun sistem QA terdiri atas 3 komponen yaitu penganalisis pertanyaan, pengambil paragraf, dan pencari jawaban. Dalam komponen penganalisis pertanyaan, dengan asumsi bahwa pertanyaan yang diajukan merupakan pertanyaan sederhana, digunakan sistem yang berbasis aturan sederhana dengan mengandalkan kata pertanyaan yang digunakan (“apa”, “mengapa”, dan “bagaimana”). Paragraf diperoleh dengan menggunakan pencarian kata kunci baik dengan menggunakan stemming ataupun tidak. Untuk pencari jawaban, jawaban diperoleh dengan menggunakan pola kata-kata khusus yang ditetapkan sebelumnya untuk setiap jenis pertanyaan. Dalam komponen pencari jawaban ini, diperoleh kesimpulan bahwa penggunaan kata kunci non-stemmed bersamaan dengan kata kunci hasil stemming memberikan nilai akurasi jawaban yang lebih baik, jika dibandingkan dengan penggunaan kata kunci non-stemmed saja atau kata kunci stem saja. Dengan menggunakan 90 pertanyaan yang dikumpulkan dari 10 orang Indonesia dan 61 dokumen sumber, diperoleh nilai MRR 0.7689, 0.5925, dan 0.5704 untuk tipe pertanyaan definisi, alasan, dan metode secara berurutan.

Focus of this research is to develop QA data and system in Bahasa Indonesia for non-factoid questions. This research is the first non-factoid QA for Bahasa Indonesia. QA system consists of three components: question analyzer, paragraph taker, and answer seeker. In the component of question analyzer, by assuming that the question posed is a simple question, we used a simple rule-based system by relying on the question word used (“what”, “why”, and “how”). On the components of paragraph taker, the paragraph is obtained by using keyword, either by using stemming or not. For answer seeker, the answers obtained by using specific word patterns that previously defined for each type of question. In the component of answer seeker, the conclusion is the use of non-stemmed keywords in conjunction with the keyword stemming results give a better answer accuracy compared to non-use of the keyword or keywords are stemmed stem only. By using 90 questions, we collected from 10 people of Indonesia and the 61 source documents, obtained MRR values 0.7689, 0.5925, and 0.5704 for type definition question, reason, and methods respectively.