ABSTRACT

ADI DARLIANSYAH. Indonesian Temporal Question Answering System. Supervised by AHMAD RIDHA.

Time is an important dimension in information retrieval. Temporal expressions describe time information embedded in the documents. Therefore, extraction and normalization of temporal expressions from documents are crucial. In this research, a question answering system is implemented for temporal information processing from documents in Indonesian language based on four types of temporal question beginning with question words such as siapa (what), kapan (when), di mana (where), and berapa (how many). Implicit time references in document are first normalized and tagged manually into explicit time references. Complex temporal question is divided into simpler questions by using temporal signal detection for specific sequence of events. In order to obtain answer candidates, heuristic weighting is performed on the top passages. Answer extraction is performed using the smallest distance between query and answer candidates. A corpus containing 100 documents and 80 queries is used in this research. Answer evaluation is based on three criteria, namely, Right, Wrong, and Unsupported. The questions are used to evaluate the results of BM25 and Proximity ranking modes. The evaluation for simple temporal questions (Type 1 and 2) using BM25 and Proximity gave the same results at 85% Right answers for Type 1 and 75% for Type 2. The results for complex temporal questions (Type 3 and 4) indicated good performance. The best results were obtained by BM25 at 95% Right answers for Type 3 and 75% for Type 4, while using Proximity resulted in 85% Right answers for Type 3 and 80% for Type 4. We also used our corpus on a non-temporal question answering system by Umriadi in 2011. The results are 60%, 55%, 60%, and 40% Right answers for Type 1, 2, 3, and 4, respectively, much lower than our temporal question answering system. Therefore, temporal expression extraction and temporal signal identification are particularly important for handling questions containing temporal information. Our system is able to identify and answer the temporal questions in Indonesian language.

Keywords: bahasa Indonesia, question answering, temporal expression, temporal question