

ABSTRAK

Ekstraksi template merupakan salah satu bentuk Ekstraksi Informasi (IE) yang bertujuan untuk memperoleh pola dari data. Beberapa penelitian menganggap bahwa ekstraksi templat bermanfaat dalam beberapa tugas seperti ekstraksi templat pertanyaan. Namun, pemanfaatan pendekatan pelabelan urutan dalam ekstraksi template pertanyaan sebagai tugas kurang dieksplorasi dalam penelitian sebelumnya. Beberapa penelitian menunjukkan bahwa pelabelan urutan bermanfaat di IE dan dapat diterapkan dalam ekstraksi templat. Namun, dalam penelitian tersebut, penjelasan rinci tentang penggunaan pendekatan pelabelan urutan dalam tugas ekstraksi template pertanyaan belum dapat dibenarkan dan eksplorasi dan analisis lebih lanjut tentang ekstraksi template pertanyaan belum dilakukan. Berdasarkan keterbatasan tersebut, penelitian ini mengusulkan model ekstraksi templat pertanyaan baru menggunakan pendekatan pelabelan sekuens dengan memanfaatkan algoritma BiLSTM. Berdasarkan hasil eksperimen dan dilanjutkan dengan analisis komprehensif yang dilakukan dalam penelitian ini, dapat disimpulkan bahwa model yang diusulkan lebih unggul dari model state-of-the-art dari penelitian terkait sebelumnya dalam mengekstraksi template pertanyaan menggunakan pendekatan sequence labeling. Model dengan performa terbaik dalam penelitian ini adalah BiLSTM yang memperoleh skor ROUGE-L sebesar 0,857. Model ini menggunakan 200 unit LSTM dan menggunakan Adam sebagai pengoptimal dengan kecepatan pembelajaran 0,001 dan dilatih menggunakan 100 epoch.

Kata kunci: NLP, ekstraksi informasi, BiLSTM, CRF, pelabelan urutan