

Deteksi Parafrasa Bahasa Indonesia menggunakan Bidirectional Long Short-Term Memory dan Bidirectional Gated Recurrent Unit = Indonesian Paraphrase Detection using Bidirectional Long Short-Term Memory And Bidirectional Gated Recurrent Unit

Stanley Pratama, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20524475&lokasi=lokal>

Abstrak

Parafrasa merupakan suatu cara untuk menuliskan kalimat dengan kata-kata lain dengan maksud atau tujuan yang sama. Pendeteksian parafrasa otomatis dapat dilakukan dengan menggunakan Natural Language Sentence Matching (NLSM) yang merupakan bagian dari Natural Language Processing (NLP). NLP merupakan teknik komputasi untuk memproses teks secara umum, sedangkan NLSM dikhususkan untuk mencari hubungan antar dua kalimat. Dengan adanya perkembangan neural network (NN), maka saat ini NLP dapat lebih mudah dilakukan oleh komputer. Model untuk mendeteksi maupun membuat parafrasa Bahasa Inggris sudah banyak dikembangkan dibandingkan dengan Bahasa Indonesia yang data pelatihannya lebih sedikit. Penelitian ini mengusulkan Model SPratama yang memodelkan deteksi parafrasa untuk Bahasa Indonesia menggunakan recurrent neural network (RNN) yaitu bidirectional long short-term memory (BiLSTM) dan bidirectional gated recurrent unit (BiGRU). Data yang digunakan adalah "Quora Question Pairs" yang diambil dari Kaggle dan diterjemahkan ke Bahasa Indonesia menggunakan Google Translate. Hasil penelitian ini menunjukkan bahwa model-model yang diusulkan mendapatkan akurasi sekitar 80% untuk pendeteksian kalimat parafrasa.

.....Paraphrasing is a way to write sentences with other words with the same intent or purpose. Automatic paraphrase detection can be done using Natural Language Sentence Matching (NLSM) which is part of Natural Language Processing (NLP). NLP is a computational technique for processing text in general, while NLSM is used specifically to find the relationship between two sentences. With the development neural network (NN), nowadays NLP can be done more easily by computers. Many models for detecting and paraphrasing in English have been developed compared to Indonesian, which has less training data. This study proposes SPratamaModel, which models paraphrase detection for Indonesian using a recurrent neural network (RNN), namely bidirectional long short-term memory (BiLSTM) and bidirectional gated recurrent unit (BiGRU). The data used is "Quora Question Pairs" taken from Kaggle and translated into Indonesian using Google Translate. The results of this study indicate that the proposed models have the accuracy of around 80% for the detection of paraphrased sentences.