

Aplikasi Bidirectional Encoder Representations from Transformers untuk Pemeringkatan Teks Bahasa Indonesia = Bidirectional Encoder Representations from Transformers Application for Text Ranking in Indonesian

Carles Octavianus, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920542859&lokasi=lokal>

Abstrak

Peningkatan jumlah data teks digital membuat manusia membutuhkan mekanisme untuk mengembalikan teks yang efektif dan efisien. Salah satu mekanisme untuk mengembalikan teks adalah dengan pemeringkatan teks. Tujuan dari pemeringkatan teks adalah menghasilkan daftar teks yang terurut berdasarkan relevansinya dalam menanggapi permintaan kueri pengguna. Pada penelitian ini, penulis menggunakan Bidirectional Encoder Representations from Transformers (BERT) untuk membangun model pemeringkatan teks berbahasa Indonesia. Terdapat 2 cara penggunaan BERT untuk pemeringkatan teks, yaitu BERT untuk klasifikasi relevansi dan BERT untuk menghasilkan representasi vektor dari teks. Pada penelitian ini, 2 cara penggunaan BERT tersebut terbagi menjadi 4 model, yaitu BERTCAT, BERTDOT, BERTDOTHardnegs, BERTDOTKD. Penggunaan BERT memberikan peningkatan kualitas pemeringkatan teks bila dibandingkan dengan model baseline BM25. Peningkatan kualitas pemeringkatan teks tersebut dapat dilihat dari nilai metrik reciprocal rank (RR), recall (R), dan normalized discounted cumulative gain (nDCG).

.....The increase in the amount of digital text data has led humans to require mechanisms for effectively and efficiently retrieving text. One mechanism for text retrieval is text ranking. The goal of text ranking is to generate a list of texts sorted based on their relevance in response to user query requests. In this study, the author uses Bidirectional Encoder Representations from Transformers (BERT) to build a text ranking model for the Indonesian language. There are 2 ways to use BERT for text ranking, namely BERT for relevance classification and BERT for generating vector representations of text. In this study, these 2 ways of using BERT are divided into 4 models, namely BERTCAT, BERTDOT, BERTDOTHardnegs, BERTDOTKD. The use of BERT improves the quality of text ranking compared to the baseline BM25 model. The improvement in the quality of text ranking can be seen from the values of the reciprocal rank (RR), recall (R), and normalized discounted cumulative gain (nDCG) metrics.