

Pencarian Dosen Pakar Menggunakan Pre-Trained Language Model BERT = Academic Expert Finding Using BERT Pre-Trained Language Model

Ilma Alpha Mannix, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920540924&lokasi=lokal>

Abstrak

Penelitian ini bertujuan untuk menguji efektivitas pre-trained language model BERT pada tugas pencarian dosen pakar. Bidirectional Encoder Representations from Transformers (BERT) merupakan salah satu state-of-the-art model saat ini yang menerapkan contextual word representation (contextual embedding). Dataset yang digunakan pada penelitian ini terdiri dari data pakar dan bukti kepakaran. Data pakar merupakan data dosen Fakultas Ilmu Komputer Universitas Indonesia (Fasilkom UI). Data bukti kepakaran merupakan data abstrak digital tugas akhir mahasiswa Fasilkom UI. Model yang diusulkan pada penelitian ini terdiri dari tiga variasi BERT, yaitu IndoBERT (Indonesian BERT), mBERT (Multilingual BERT), dan SciBERT (Scientific BERT) yang akan dibandingkan dengan model baseline menggunakan word2vec. Terdapat dua pendekatan yang dilakukan untuk mendapatkan urutan dosen pakar pada variasi model BERT, yaitu pendekatan feature-based dan fine-tuning. Penelitian ini menunjukkan bahwa model IndoBERT dengan pendekatan feature-based memberikan hasil yang lebih baik dibandingkan baseline dengan peningkatan 6% untuk metrik MRR hingga 9% untuk metrik NDCG@10. Pendekatan fine-tuning juga memberikan hasil yang lebih baik pada model IndoBERT dibandingkan baseline dengan peningkatan 10% untuk metrik MRR hingga 18% untuk metrik P@5. Diantara kedua pendekatan tersebut, dibuktikan bahwa pendekatan fine-tuning memberikan hasil yang lebih baik dibandingkan dengan pendekatan feature-based dengan peningkatan 1% untuk metrik P@10 hingga 5% untuk metrik MRR. Penelitian ini menunjukkan bahwa penggunaan pre-trained language model BERT memberikan hasil yang lebih baik dibandingkan baseline word2vec dalam tugas pencarian dosen pakar.

.....This study aims to test the effectiveness of the pre-trained language model BERT on the task of expert finding. Bidirectional Encoder Representations from Transformers (BERT) is one of the current state-of-the-art models that applies contextual word representation (contextual embedding). The dataset used in this study consists of expert data and expertise evidence. The expert data is composed of faculty members from the Faculty of Computer Science, University of Indonesia (Fasilkom UI). The expertise evidence data consists of digital abstracts by Fasilkom UI students. The proposed model in this research consists of three variations of BERT, namely IndoBERT (Indonesian BERT), mBERT (Multilingual BERT), and SciBERT (Scientific BERT), which will be compared to a baseline model using word2vec. Two approaches were employed to obtain the ranking of expert faculty members using the BERT variations, namely the feature-based approach and fine-tuning. The results of this study shows that the IndoBERT model with the feature-based approach outperforms the baseline, with an improvement of 6% for the MRR metric and up to 9% for the NDCG@10 metric. The fine-tuning approach also yields better results for the IndoBERT model compared to the baseline, with an improvement of 10% for the MRR metric and up to 18% for the P@5 metric. Among these two approaches, it is proven that the fine-tuning approach performs better than the feature-based approach, with an improvement of 1% for the P@10 metric and up to 5% for the MRR metric. This research shows that the use of the pre-trained language model BERT provides better results compared

to the baseline word2vec in the task of expert finding.