

Pengenalan Pertanyaan, Entitas, Dan Frase Kunci Pada Data Forum Tanya Jawab Kesehatan Menggunakan BERT = Identification Of Questions, Entities, And Keyphrases From Consumer Health Forum Using BERT

Tsaqif Naufal, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920519495&lokasi=lokal>

Abstrak

Seiring dengan semakin banyaknya masyarakat yang menggunakan forum tanya-jawab kesehatan online, kebutuhan akan adanya sistem tanya-jawab kesehatan yang dapat berjalan secara otomatis semakin besar. Salah satu bagian penting dari sistem tanya-jawab kesehatan otomatis adalah question processing untuk mendapatkan informasi relevan dari pertanyaan pengguna. Terdapat beberapa task yang merupakan bagian dari question processing, di antaranya pengenalan pertanyaan, pengenalan entitas kesehatan, dan ekstraksi frase kunci. Pada penelitian ini, penulis mencoba tiga model untuk menyelesaikan ketiga task tersebut, yaitu IndoDistilBERT, IndoDistilBERT-BiLSTMs, dan IndoDistilBERT-BiLSTMs-CRF. Hasil eksperimen menunjukkan bahwa IndoDistilBERT-BiLSTMs-CRF memberikan hasil terbaik untuk task pengenalan pertanyaan dengan skor F1 sebesar 94,45%, lebih baik 3,15% dibandingkan baseline. Untuk task pengenalan entitas kesehatan, IndoDistilBERT-BiLSTMs memberikan hasil terbaik dengan skor F1 sebesar 73,78%, lebih baik 3,53% dibandingkan baseline. Untuk task ekstraksi frase kunci, model IndoDistilBERT-BiLSTMs memberikan hasil terbaik dengan skor F1 sebesar 77,42%, lebih baik 4,25% dibandingkan baseline. Selain itu, percobaan dengan pendekatan multi-task learning untuk menyelesaikan task pengenalan entitas kesehatan dan ekstraksi frase kunci belum mampu mengungguli hasil dari pendekatan single-task learning untuk masing-masing task.

.....With the increasing number of people who use health question-and-answer online forum, the need for a health question-and-answer system that can run automatically is getting bigger. One of the important parts of an automated health question-and-answer system is question processing to get relevant information from user queries. There are several tasks which are part of question processing, including question recognition, medical entity recognition, and keyphrases extraction. On this research, we try three models to solve those three tasks, namely IndoDistilBERT, IndoDistilBERT-BiLSTMs, and IndoDistilBERT-BiLSTMs-CRF. Our experiment shows that IndoDistilBERT-BiLSTMs-CRF gives the best results for question recognition task with F1-score of 94,45%, 3,15% better than baseline. For medical entity recognition task, IndoDistilBERT-BiLSTMs gives the best results with F1-score of 73,78%, 3,53% better than baseline. For keyphrases extraction task, IndoDistilBERT-BiLSTMs gives the best results with F1-score of 77,42%, 4,25% better than baseline. Besides that, experiments with multi-task learning approach to solve medical entity recognition and keyphrases extraction have not been able to outperform the results of single-task learning approach for each task.