

Deteksi Pertanyaan Duplikat Pada Forum Kesehatan Berbahasa Indonesia Dengan Ekspansi Kata = Duplicate Question Detection in Indonesian Health Forum With the Help of Term Expansion

Heidi Renata Halim, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551863&lokasi=lokal>

Abstrak

Seiring dengan majunya teknologi di Indonesia, banyak layanan kesehatan online yang bermunculan. Pengguna bisa bertanya langsung pada tenaga medis profesional tiap mereka memiliki masalah kesehatan ringan yang tidak membutuhkan janji temu langsung dengan dokter. Sebagai pengguna, tentunya mereka mengharapkan respon yang cepat dari situs yang mereka gunakan, hal ini kedengarannya mustahil dilakukan karena tidak semua tenaga medis profesional yang bekerja pada layanan medis tersebut ada setiap saat memantau semua pertanyaan yang masuk. Namun, hal ini bisa dilakukan dengan cara mencocokkan pertanyaan yang baru dimasukkan dan mencari pertanyaan yang sudah pernah ditanyakan di masa lalu yang memiliki persamaan dengan pertanyaan yang baru dimasukkan. Secara singkat, kita bisa mencari duplikat dari pertanyaan yang ditanyakan oleh pengguna dan mengembalikan jawaban dari pertanyaan duplikat tersebut daripada menunggu jawaban langsung dari dokter. Penelitian ini akan menggunakan pendekatan temu balik informasi dalam mendeteksi pertanyaan duplikat yang pernah ditanyakan di masa lalu. Selain itu, penelitian ini juga akan mengkombinasikan ekspansi kata yang dilakukan kepada kueri, dokumen, serta filter kata-kata stopwords untuk meningkatkan skor reciprocal-rank dan recall dari model yang digunakan. Hasil penelitian ini menyimpulkan bahwa ekspansi kata yang dilakukan pada kueri serta dokumen tidak menghasilkan skor reciprocal rank dan recall yang lebih baik. Penggunaan word embedding untuk memperbanyak kata stopwords yang dihapus dari data mampu menghasilkan skor reciprocal rank yang lebih tinggi meskipun nilainya belum signifikan.

.....With the advancement of technology and internet in Indonesia, many online healthcare services have emerged where users can directly consult with medical professionals if they have minor health issues that do not require an in-person appointment with a doctor. As users, they naturally expect quick responses from the sites they use. This seems impossible to do as not all medical professionals working who are working on these services are always available to monitor every incoming question. However, this can be achieved by matching newly submitted questions with previously asked questions that have similarities. In short, we can search for duplicates of the questions asked by users and return answers from those duplicate questions instead of waiting for a direct response from a doctor. This research will use an information retrieval approach to detect duplicate questions that have been asked in the past. Additionally, this study will combine query expansion, document expansion, and stopwords filtering to improve the reciprocal-rank and recall scores of the model used. This research concludes that query and document expansion do not yield better reciprocal rank and recall scores. On the other hand, using word embedding to expand the stopwords list removed from the data can help achieve higher reciprocal rank scores, although the improvement displays are still not significant enough to be categorized as a major change.