

Pemeringkatan Dokumen Secara Temporal dengan Dynamic Embeddings = Temporal Ranking with Dynamic Embeddings

Dinda Sigmawaty, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20529062&lokasi=lokal>

Abstrak

Saat mencari artikel yang diterbitkan dalam periode waktu yang panjang, pengguna biasanya membutuhkan dokumen yang tidak hanya relevan terhadap topik tetapi juga relevan terhadap waktu. Tesis ini membahas tentang pemeringkatan dokumen dengan konsep waktu atau temporal, di mana dokumen dengan topik dan waktu yang dekat dengan query harus diberikan peringkat yang lebih tinggi. Untuk mengetahui waktu yang sesuai dengan query pengguna, tesis ini mengembangkan teknik pemeringkatan temporal yang diperoleh dari distribusi keterkaitan kata dari waktu ke waktu yang dipelajari pada sebuah arsip berita dalam Bahasa Indonesia. Keterkaitan kata dipelajari menggunakan Dynamic Embeddings yaitu Word2Vec yang dipelajari terpisah dari waktu ke waktu, OrthoTrans-Word2Vec dan Dynamic Bernoulli Embeddings. Dalam menangkap relevansi secara topikal, model yang diusulkan menggunakan Dual Embedding Space Model (DESM) yang dibangun dengan teknik temporal sesuai dengan waktu pembuatan dokumen. Untuk meningkatkan nilai presisi, model tersebut juga menggunakan sebuah klasifikasi temporal yang dipelajari menggunakan Support Vector Machine (SVM) dan Basis Threshold. Skor tertinggi dicapai ketika membangun model menggunakan Word2Vec yaitu 66% pada presisi rata-rata dan 68% pada presisi awal. Model tersebut juga terbukti efektif pada query temporal yang memiliki pola seperti tren, periodisitas, dan musiman.

.....When searching for articles published over time, users usually require documents that are not only topically relevant but also created during relevant time periods. This thesis studied about document ranking with temporal concept, where documents with topic and time that closely match with the queries should be ranking higher. In order to capturing the time of user query intent, the models developed with temporal ranking technique from distribution of word relatedness over time learned from news archive in Bahasa Indonesia. Word relatedness captured by using Dynamic embeddings, such as Word2Vec learned separately over time, OrthoTrans-Word2Vec dan Dynamic Bernoulli Embeddings. For capturing topical relevance, the proposed model used Dual Embedding Space Model (DESM) in the temporal technique according to document timestamp. The model also combined with temporal classification using Support Vector Machine (SVM) and threshold-based strategy. The highest score was achieved by a model using Word2Vec, which is 66% in average precision and 68% in early precision. The result also showed that the model is effective in capturing temporal patterns such as spikes, periodicity, and seasonality