Pengembangan sistem pencarian pakar dengan query expansion menggunakan Word Embedding, Document Embedding, dan Ontologi = Developing Expert Search System with Query Expansion Using Word Embedding, Document Embedding, and Ontology

Theresia Veronika Rampisela, author

Abstrak

Penelitian ini mencari dosen pakar di Fakultas Ilmu Komputer Universitas Indonesia (Fasilkom UI) dengan menggunakan data abstrak dan metadata tugas akhir mahasiswa Fasilkom UI menggunakan teknik information retrieval. Pencarian pakar dilakukan tanpa query expansion dan dengan query expansion. Metode yang digunakan untuk mencari dosen pakar adalah metode berbasis BM25 serta kombinasi antara word2vec dan doc2vec, yang merupakan word embedding dan document embedding. Teknik yang digunakan untuk mengatasi masalah vocabulary mismatch adalah teknik query expansion dengan pendekatan statistik, semantik, serta hybrid. Kontribusi penelitian ini adalah 2 metode baru untuk pencarian pakar tanpa query expansion, 6 jenis kombinasi relasi ontologi berdasarkan concept hierarchy ACM CCS 2012 untuk mengekspansi kueri, serta 14 jenis kombinasi antara metode query expansion berbasis embedding dan ontologi. Evaluasi dari hasil pencarian pakar dilakukan dengan menghitung binary relevance berdasarkan human judgment. Pada penelitian ini, metode pencarian pakar tanpa query expansion dengan hasil yang paling baik adalah metode BM25-sum dengan nilai AP@5 sebesar 0.648. Pencarian pakar dengan berbagai macam jenis query expansion tidak dapat meningkatkan performa retrieval tanpa query expansion secara signifikan, tetapi pencarian dengan query expansion menggunakan metode berbasis embedding, yaitu w2v-w2v memiliki nilai AP@5 sebesar 0.696, yang lebih tinggi dari skor AP@5 dari BM25-sum tanpa query expansion. Pencarian pakar dengan query expansion berbasis ontologi serta hybrid mampu menghasilkan skor AP@5 setinggi 0.664. Skor tersebut tidak setinggi pencarian dengan query expansion dengan w2v-w2v, namun hasilnya lebih baik dari pencarian tanpa query expansion. <hr /> This research searches for expert lecturers in the Faculty of Computer Science, Universitas Indonesia (Fasilkom UI) with information retrieval techniques using students’ thesis abstract and metadata. The retrieval process is done without and with query expansion. The methods used to find expert lecturers are BM25-based methods as well as combinations between word2vec and doc2vec, which are word embedding and document embedding. We performed query expansion using statistical, semantic, and hybrid approaches to solve vocabulary mismatch problems. This research’s contributions are 2 new methods to retrieve experts without query expansion, 6 types of ontological relations based on the ACM CCS 2012 concept hierarchy to expand queries, and 14 types of combinations between embedding-based and ontology-based query expansion methods. The expert retrieval result is evaluated by calculating binary relevance based on human judgment. Expert search method without query expansion that produces the best result in this research is the BM25-sum method, with an AP@5 score of 0.648. Even though expert retrieval with various query expansion methods does not increase the performance of retrieval without query expansion significantly, the expert

search method with embedding-based query expansion method, i.e. w2v-w2v, achieved an AP@5 score of 0.696, which is higher than that of BM25-sum without query expansion. Ontology-based and hybrid query expansion expert search methods managed to score 0.664 for AP@5. This score is not as high as that of w2v-w2v, but the result is still better than that of retrieval without query expansion.</p>