ABSTRAK

Arsitektur *content-based Publish-Subscribe* memungkinkan pencocokan pesan dengan *subscripiton* berdasarkan isi pesan yang diterima. Namun, pada lingkungan yang menghasilkan aliran teks pendek seperti media sosial, proses pencocokan dapat terkendala karena keterbatasan bukti tekstual dan variasi bahasa alami yang menyebabkan banyak pesan relevan tidak terdeteksi. Penelitian ini mengusulkan penerapan kesadaran semantik (*semantic awareness*) melalui teknik perluasan kueri menggunakan *Global Term Relationship Model* (GTRM) untuk meningkatkan efektivitas *event matching* pada arsitektur Publish-Subscribe. Penelitian menggunakan dataset TREC-Microblog yang diadaptasi untuk lingkungan Publish-Subscribe, di mana topik kueri diperlakukan sebagai *subscription* dan tweet sebagai *message event* Evaluasi dilakukan terhadap enam strategi pencocokan, yaitu *Loose OR*, *Strict AND*, *Loose OR with GTRM*, *Strict AND with GTRM*, *Threshold-based Matching with GTRM*, dan *Dynamic Threshold with GTRM*. Kinerja pendekatan diukur menggunakan metrik *Precision*, *Recall*, dant *F1-Score*. Hasil simulai menunjukkan bahwa penerapan ekspansi dengan GTRM mampu meningkatkan kemampuan sistem dalam menemukan pesan yang relevan dibandingkan pencocokan kata kunci murni. Strategi yang menggunakan GTRM secara konsisten menghasilkan nilai *recall* yang lebih tinggi, sedangkan *Threshold-based Matching with GTRM* memberikan keseimbangan yang lebih baik antara precision dan recall. Selain itu, mekanisme *Dynamic Threshold* terbukti mampu meningkatkan jumlah pesan relevan yang diterima pada topik-topik yang mengandung bukti tekstual yang terbatas (*sparse-event*) tanpa memerlukan penyesuaian *threshold* secara manual.

Kata kunci: *Publish-Subscribe, Event Matching, Query Expansion, Semantic Awareness, Global Term Relationship Model, Information Retrieval, TREC-Microblog*

ABSTRACT

Content-based Publish-Subscribe architecture enables message matching with subscriptions based on the content of the received messages. However, in environments that generate streams of short text, such as social media, the matching process can be hindered by limited textual evidence and variations in natural language, resulting in many relevant messages going undetected. This research proposes the application of semantic awareness through query expansion techniques using the Global Term Relationship Model (GTRM) to improve the effectiveness of event matching in the Publish-Subscribe architecture. The study uses the TREC-Microblog dataset adapted for the Publish-Subscribe environment, where query topics are treated as subscriptions and tweets as message events. The evaluation was conducted on six matching strategies: Loose OR, Strict AND, Loose OR with GTRM, Strict AND with GTRM, Threshold-based Matching with GTRM, and Dynamic Threshold with GTRM. The performance of the approaches was measured using the Precision, Recall, and F1-Score metrics. Simulation results show that the application of expansion with GTRM is able to improve the system's ability to find relevant messages compared to pure keyword matching. Strategies using GTRM consistently yield higher recall values, while Threshold-based Matching with GTRM provides a better balance between precision and recall. Additionally, the Dynamic Threshold mechanism has proven capable of increasing the number of relevant messages received on topics containing limited textual evidence (sparse-event) without requiring manual threshold adjustment.

Keywords: Publish-Subscribe, Event Matching, Query Expansion, Semantic Awareness, Global Term Relationship Model, Information Retrieval, TREC-Microblog