

1. Pendahuluan

Pencarian informasi di internet sudah menjadi kebutuhan bagi sebagian besar orang dengan berbagai tujuan, baik untuk mencari berita terkini, artikel ilmiah, maupun hiburan. Karena sudah menjadi kebutuhan, manusia dituntut untuk dapat memenuhi kebutuhan tersebut, salah satunya adalah dengan mencarinya di internet [1]. Untuk pencarian artikel ilmiah sendiri bisa menggunakan mesin pencari *google scholar* untuk memudahkan pencarian agar lebih spesifik untuk artikel ilmiah. Meski menggunakan mesin pencari *google scholar*, informasi yang disediakan masih terbilang banyak dan membutuhkan kata kunci yang tepat untuk mendapatkan informasi yang sesuai dan relevan.

Ekspansi query adalah salah satu teknik dari *query reformulation* yang dilakukan dengan menambahkan term kedalamnya [3]. Untuk penelitian ini term yang ditambahkan adalah kata-kata yang memiliki kemiripan arti yang didapat dari hasil *word embedding* menggunakan *word2vec*. Ekspansi query berkaitan erat dengan *information retrieval* atau IR. Karena ekspansi query memiliki peran penting dalam proses melakukan *information retrieval* [4]. Dengan melakukan pencarian pada mesin pencari *google scholar* sebagai proses *information retrieval*, maka penelitian mengenai ekspansi query dengan *word2vec* ini dapat dilaksanakan.

Proses ekspansi query sendiri dapat menggunakan berbagai metode, salah satunya *word embedding*. Bahkan pada penelitian yang dilakukan oleh Alfredo Silva dan Marcelo Mendoza dimana mereka meneliti bagaimana metode *word embedding* dapat meningkatkan performa dari ekspansi query [4]. Pada penelitian tersebut dilakukan percobaan ekspansi query dengan menggunakan pendekatan *inverse document frequency* dan *average word embedding* atau IDF-AWE, sebuah representasi vector untuk query berbasis AQE (*automated query expansion*) lalu ditambahkan dengan berbagai metode *word embedding* seperti *glove*, *word2vec*, *FastText*, dll [4]. Hasil pada penelitian tersebut menunjukkan bahwa metode *word embedding* memberikan hasil yang lebih baik ketika digunakan bersama AQE dan IDF-AWE. Lalu penelitian yang dilakukan Claudio dan Giovanni memberi penjelasan lebih lanjut mengenai penerapan IR seperti *interactive query refinement*, *relevance feedback*, *word sense disambiguation in IR*, dan *search result clustering* [6]. Beserta penerapan lainnya pada penggunaan *automated query expansion* atau AQE seperti *question answering*, *multimedia information retrieval*, *information filtering*, dan *cross-language information retrieval* [6]. Penelitian tersebut memberikan *system ranking* untuk menjadi *variable* pembanding untuk dapat mencari nilai *similarity* antara query dan dokumen terkait.

Kebanyakan dari penelitian mengenai ekspansi query menggunakan AQE sebagai basis awalnya, selanjutnya menambahkan proses lain seperti *word embedding* dan sebagainya. Seperti pada penelitian yang dilakukan Saar Kuzi dan rekannya dimana mereka menggunakan *word2vec* untuk melakukan ekspansi query [7]. Pada penelitian ini, akan menggunakan *word2vec* untuk ekspansi query, dan menggunakan mesin pencari *google scholar* untuk melakukan dua skenario pencarian untuk proses *information retrieval*. selanjutnya melakukan survey pada lima orang partisipan untuk menilai tingkat relevansi pada hasil skenario pencarian untuk *relevance feedback* yang digunakan sebagai metode penilaian dari kualitas query. Dataset yang digunakan diambil dari *repository WING-NUS/scisumm-corpus* yang dimana terdapat 1000 set dokumen artikel ilmiah yang sudah teranotasi dari *scisummnet* [8].

Tujuan penelitian ini yaitu untuk mengetahui performansi dari penggunaan ekspansi query dengan menggunakan *word2vec* dalam melakukan pencarian artikel ilmiah di internet.