

Abstrak

Di era sekarang ini kebutuhan informasi semakin tinggi dengan banyaknya teknologi yang berkembang dengan cepat. Dengan adanya internet yang semakin cepat dan efektif, maka sistem *Community Question Answering* (CQA) sudah dapat dipastikan akan sangat membantu pengguna internet untuk mendapatkan informasi yang dibutuhkan. Dengan masukan berupa dataset berbentuk XML yang berisi pertanyaan baru, pertanyaan relevan dan jawaban. *Output* yang dihasilkan berupa nilai *Mean Average Precision* (MAP) dari sepuluh komentar *good* teratas..

CQA sendiri cukup terbuka untuk umum dan semuanya bebas untuk bertanyajawab, tetapi dengan kebebasan itu pengguna juga disulitkan dengan banyaknya jawaban dan tidak menjamin semuanya benar dan sesuai dengan pertanyaan. Bahkan ada kemungkinan juga jika jawaban yang terbaik ada di pertanyaan lain yang sudah pernah ditanyakan.

Penelitian yang ada sebelumnya menggunakan fitur *Cosine Similarity*. Fitur *Cosine Similarity* hanya mengambil jawaban yang memiliki kesamaan kata dengan pertanyaan yang ada. Sedangkan dengan ditambahkan fitur *Soft-Cosine Semantic Similarity* akan meningkatkan kemungkinan untuk mendapatkan jawaban yang tepat meskipun tidak memiliki kesamaan kata sekalipun.

Pengujian dilakukan menggunakan *dataset* dari SemEval-2017 Task 3 menunjukkan bahwa gabungan fitur *Soft-Cosine Semantic Similarity* dengan algoritma klasifikasi *Support Vector Machine* lebih baik dari kombinasi yang lain. Kombinasi ini menghasilkan nilai MAP sebesar **21.0%** untuk mencari persamaan *Original Question* dengan *Related Comments*.

Kata Kunci : *community question answering*, klasifikasi, pemberian peringkat komentar, *qatar living forum*, *soft-cosine semantic similarity*
