

Abstrak

Penelitian ini membahas tentang analisis dan implementasi suatu sistem yang mampu secara otomatis menjawab pertanyaan yang dibutuhkan user berdasarkan FAQ yang telah ada dengan mengukur *similarity* diantara keduanya. Selanjutnya jawaban dari pertanyaan pada data FAQ yang memiliki nilai kecocokan paling tinggi diasumsikan dapat menjawab pertanyaan *user*. Pengujian dilakukan menggunakan *dataset* SemEval-2016 Task 3. Dalam prosesnya, pertanyaan user dan data FAQ akan melewati proses preprocessing yang menghasilkan token-token dan selanjutnya dilakukan pengukuran *similarity* menggunakan *overall similarity*. *Overall similarity* merupakan hasil kombinasi linear nilai kecocokan dari *semantic similarity* dan *statistic similarity*.

Hasil pengujian beberapa skenario menunjukkan bahwa nilai MAP paling tinggi diperoleh 59,65%, dengan hasil pengujian lainnya masing-masing precision yang diperoleh sebesar 44%, nilai recall 100%, nilai F1-measure sebesar 61,11%, dan nilai Accuracy sebesar 71,43%.

Kata kunci: *similarity, FAQ, overall similarity, semantic, accuracy, precision and recall.*