

TELKOM UNIVERISTY

Abstrak

Fakultas Informatika

Sekolah Pascasarjana Teknik Informatika

Magister Informatika

Answer Selection Using Word Alignment Based On Part Of Speech Tagging In**Community Question Answering**

by Ade Sutedi

Community Question-Answering (CQA) merupakan forum *online* dimana pengguna bisa mengajukan pertanyaan dan pengguna lainnya memberikan komentar terkait permasalahan yang ditanyakan. Oleh karena CQA tidak ada batasan dalam menyampaikan pertanyaan atau jawaban, maka banyak komentar yang tidak sesuai dengan masalah yang diajukan sebelumnya. Untuk mengatasi masalah tersebut, dalam penelitian sebelumnya digunakan pendekatan dengan menggabungkan fitur-fitur *lexical* dan *semantic similarity*. Hasil yang diperoleh menunjukkan bahwa pendekatan tersebut hanya cocok untuk masalah *similarity* saja dan tidak cocok untuk masalah CQA. Berdasarkan penelitian sebelumnya, terdapat beberapa hal yang dapat dikembangkan dan ditingkatkan. Pertama, proses pemasangan dan pencocokan kata dengan representasi vektor hanya menghitung kata yang sama persis, sehingga tidak dapat mencakup kata-kata lain yang memiliki keterkaitan dan arti yang sama. Kedua, teknik *noun overlap* untuk mengukur kemiripan kata-kata yang dipasangkan tidak menunjukkan bahwa kedua kata tersebut memiliki kesamaan arti. Oleh karena itu, jika kedua kata memiliki pola POS tag yang sama maka harus ditentukan bahwa kedua kata tersebut memiliki keterkaitan dan kesamaan. Dalam penelitian ini, metode *unsupervised lexical* dan *semantic similarity* digunakan dengan pendekatan yang berbeda dari penelitian sebelumnya dari segi kata dan kesamaan kontekstual. Data diambil dari kompetisi SemEval 2017 yang berfokus pada *Question-Answer Similarity*. Hasil eksperimen untuk *Mean Average Precision* (MAP) menunjukkan peningkatan dari 0,6742 ke 0,6845, 1,03% lebih tinggi dari penelitian sebelumnya. Peningkatan ini diperoleh dari *lexical similarity*, dimana tidak hanya pola *noun* yang digunakan tetapi juga pola *verb*. Selain itu, *semantic similarity* memiliki peran penting dalam menentukan kata-kata yang memiliki pola POS tag dan makna yang sama untuk menentukan relevansi di antara keduanya.

Kata Kunci: *Community Question-Answering, Unsupervised, Lexical* dan *Semantic Similarity*, relevansi.