

Klasifikasi Argumen Semantik Menggunakan *Head Word of Prepositional Phrase* dan *First and Last Word/ POS in Constituent*

Rani Puspitasari¹, Moch. Arif Bijaksana, Ir., M.Tech.², Angelina Prima Kurniati, ST., MT³

Fakultas Informatika Telkom University, Bandung

Abstrak

Dalam bidang linguistik, argumen semantik merupakan sebuah ekspresi yang digunakan untuk membantu mempelajari makna dari suatu predikat didalam sebuah kalimat. Penggunaan teknik yang tepat dan akurat dalam menganotasi teks yang memiliki struktur argumen semantik dapat memfasilitasi penemuan pola informasi dalam teks berukuran besar dan dapat memberikan anotasi Who, What, Whom, When, Where, Why, dan How pada kalimat dalam sebuah teks.

Penelitian tentang klasifikasi argumen semantik mengimplementasikan penggunaan metode *Support Vector Machines* (SVM) dalam *database* PropBank, menggantikan algoritma *statistical classification* yang digunakan dalam penelitian sebelumnya, kemudian mengevaluasi perubahan-perubahan yang ada untuk meningkatkan hasil performansi. Selain itu, penelitian menggunakan algoritma SVM tersebut juga memberikan fitur-fitur baru yang diharapkan dapat memberikan kemajuan performansi yang sebelumnya memiliki hasil yang kurang baik.

Tugas Akhir ini mengimplementasikan 2 fitur baru yang digunakan bersamaan dengan fitur utama baseline, yaitu *Head Word of Prepositional Phrase* dan *First and Last Word/POS in Constituent* yang kemudian diklasifikasikan menggunakan bantuan metode SVM *Sequential Minimal Optimization* (SMO). Hasil dari pengimplementasian kedua fitur ini secara bersamaan dengan fitur utama memberikan hasil akurasi rata-rata sebesar 67,99%.

Kata Kunci: *klasifikasi argumen semantik, fitur, Support Vector Machines.*

Abstract

In linguistic, semantic argument is an expression used to help understanding the meaning of a certain predicate in a sentence. The use of a precise and accurate technique to anotate a text that have semantic argument structure can facilitate the detection of information pattern inside large text dan it can also give the Who, What, Whom, When, Where, Why and How annotation of a sentence in the text.

A study about semantic argument classification implement the use of Support Vector Machine (SVM) method for Propbank database, replacing the statistical classification algorithm that is used in the previous studies, and evaluate the obvious changes for increasing the performance result. Beside that, the study using SVM method is also giving a few new features in hope that it will give an increase to the performance score that wasn't very good before.

This final project implements two new features that are going to be used together with the main feature baseline, Head Word of Prepositional Phrase along with First and Last Word/POS in Constituent and then classify it with the help of SVM Sequential Minimal Optimization (SMO) method. The result of the implementation of these two features together along with the baseline feature produce the average accuracy score of 67,99%.

Keywords : *argument semantic classification, feature, Support Vector Machines.*

1. Pendahuluan

1.1 Latar Belakang

Teknik akurat yang mencakup area permasalahan yang luas yang dapat menganotasi teks yang memiliki struktur argumen semantik dapat menghasilkan penemuan pola informasi dalam teks berukuran besar. Semantic parsing dapat memberikan anotasi dalam sebuah teks, memprediksi dan memberikan label kepada tiap predikat dalam argumen semantik [11].

Semantic parsing sendiri merupakan bagian dari natural language processing yang menentukan struktur dari sebuah kalimat, seperti contohnya kata-kata apa yang dapat digunakan dalam satu kalimat yang sama sebagai frasa, maupun kata mana yang merupakan subjek atau

objek. Proses klasifikasi argumen semantik ini melakukan klasifikasi terhadap argumen semantik hasil dari semantic parsing. Pada awalnya para peneliti menerapkan teknik machine learning yang menggunakan algoritma *statistical classification* untuk proses klasifikasi. Kemudian, sebuah penelitian tentang argumen semantik [10] mengimplementasikan penggunaan metode *Support Vector Machines* (SVM).

Support Vector Machines merupakan model supervised learning yang dapat melakukan analisa data dan pattern recognition. SVM sendiri memiliki potensi yang besar serta performansi yang sangat baik. SVM mengaplikasikan prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan yang