

ABSTRACT

Crime Information Extraction is a task to extract some entities in the crime domain. Previous researchers have studied this task using rules to extract these entities in the English dataset. However, the rules is not very precise, and make the system have miss-classification. This error is due to the inability to resolve the name entities. This study proposed a system that can extract crime-related information in Indonesian because Indonesians need to know the crime information openly based on the Crime Information Need Survey [1]. There are two main methods are implemented, as Crime Classification using Ontology and Rule-Based Crime Argument Extraction. The extraction is conducted by creating rules by combining dependency parsing and Part-Of-Speech tagging. These methods identify five crime entities: crime type, victim, perpetrator, location, and time. The evaluation is conducted by comparing the system output with the data manual labelling. The results indicate 60.70% F1-Measure, 62.43% precision, and 59.06% recall. These show that the proposed method still needs to be fixed in some areas, especially in creating a combination of rules. The system still hard to define the perpetrator entities, victim entities, and location entities.

Keywords: Crime Information Extraction, Rule-Based Information Extraction, Information Extraction