

ABSTRAK

Crime Information Extraction adalah sebuah proses untuk mengekstrak beberapa entitas pada domain kriminal. Penelitian sebelumnya sudah melakukan penelitian untuk mengekstrak entitas kriminal menggunakan aturan pada dataset berbahasa Inggris. Namun aturan-aturan tersebut tidak cukup tepat, sehingga sistem masih memiliki kesalahan pengelompokan. Kesalahan tersebut terjadi karena ketidakmampuan system dalam menemukan entitas-entitas yang dibutuhkan. Penelitian ini mengajukan sebuah sistem yang dapat mengekstrak informasi yang berkaitan dengan informasi kriminal pada Bahasa Indonesia sesuai dengan kebutuhan masyarakat Indonesia berdasarkan Survei Kebutuhan Informasi Kriminalitas di Indonesia [1]. Terdapat dua metode utama yang digunakan pada penelitian ini antara lain *Crime Classification* menggunakan *Ontology* dan *Rule-Based Crime Argument Extraction*. Sistem akan melakukan ekstraksi dengan memanfaatkan aturan yang dibangun berdasarkan kombinasi dari hasil *Dependency Parsing* dan *Part-Of-Speech Tagging*. Metode ini mengidentifikasi lima entitas kriminalitas yaitu tipe kriminalitas, korban, pelaku, tempat, dan waktu. Proses evaluasi dilakukan dengan melakukan perbandingan terhadap keluaran sistem dengan data yang telah dilabel secara manual. Hasil evaluasi yang didapatkan adalah 60,70% F1-Measure, 62.43% precision, dan 59.06% recall. Hasil ini menunjukkan bahwa sistem masih membutuhkan perbaikan di beberapa area, khususnya dalam membuat kombinasi aturan untuk proses ekstraksi entitas kriminal. Sistem juga masih kesulitan untuk menentukan entitas korban, pelaku, dan lokasi.

Kata Kunci: *Crime Information Extraction, Rule-Based Information Extraction, Information Extraction*