

# Crime event localization and deduplication

Federica Rollo<sup>[0000–0002–3834–3629]</sup> and Laura Po<sup>[0000–0002–3345–176X]</sup>

‘Enzo Ferrari’ Engineering Department  
Via Vivarelli, 10, 42125 Modena, ITALY  
`name.surname@unimore.it`

**Abstract.** Crime analysis is an approach for identifying patterns and trends in crime events, while information extraction is the task of extracting relevant information from unstructured data. If crime reports are not directly available to the public, a possible solution is to derive crime information published in newspaper articles.

This paper aims at extracting, localizing, deduplicating, and visualizing crime events from online news articles. This work demonstrates how crime-related information can be obtained from newspapers and exploited to create a consistent database of crime events with an automatic process. The approach employs a Named Entity Recognition (NER) algorithm to retrieve locations, organizations and persons and a mapping phase to link entities to Linked Data resources. The date of the event is retrieved through the temporal expressions extraction and normalization. For duplicate detection, an approach analyses and combines crime category, description, location, and crime event date to identify which news articles refer to the same event. The approach has been successfully applied in the Modena province (Italy), focusing on eleven types of crime happen from 2011 till now. The flexibility of the approach allows it to be easily adapted to other cities, regions, or countries and also to other domains.

**Keywords:** Crime analysis · Crime mapping · News extraction · Document similarity · Duplicate detection.