CEM: an Ontology for Crime Events in Newspaper Articles

Federica Rollo, Laura Po, Alessandro Castellucci
University of Modena and Reggio Emilia, Modena, IT
Contact: federica.rollo@unimore.it

38th ACM/SIGAPP SAC
Tallinn, Estonia
March 27 - 31, 2023

Motivation and goals

The adoption of semantic technologies for the representation of crime events can help Law Enforcement Agencies in crime prevention and investigation. Online newspapers and social networks are valuable sources for **crime intelligence gathering**.

Event-centric KGs allow to capture the dynamic aspects of events and represent the relationships between them, building temporal and causal chains.

Existing KGs (Wikidata, DBpedia and YAGO) focus on entity-centric information of historical events, while crime ontologies are bound to national judicial systems. For newspapers and social media, we propose **Crime Event Model**, a new lightweight ontology to model crime events that is independent of legal representation.

The **Crime Event Model** (CEM) is a new lightweight ontology to model crime events as usually described in online news articles

Instances (from existing sources)	
Classes	50 (18)
Datatype properties	10 (3)
Object properties	21 (16)

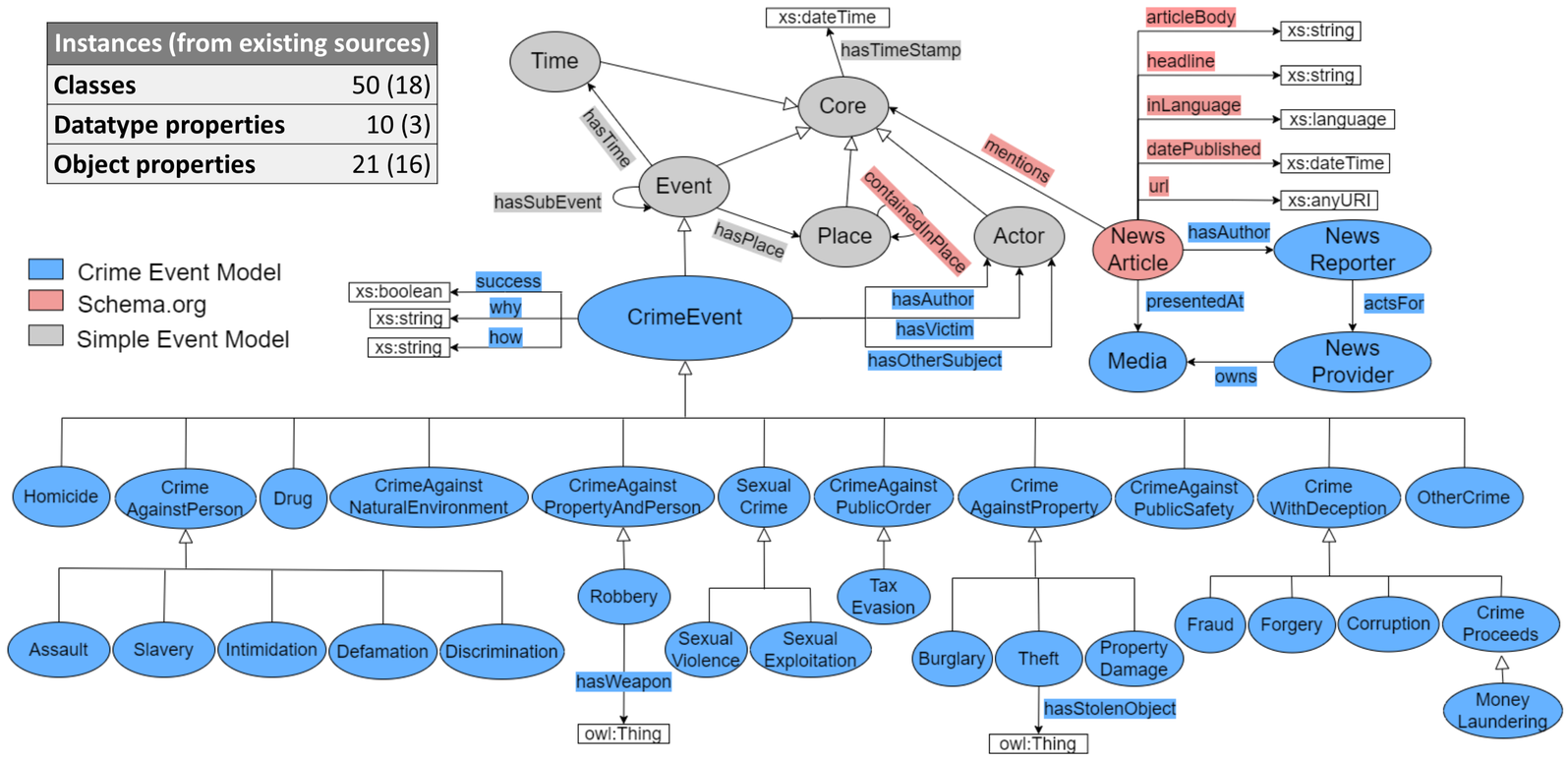

Figure 1 - Classes and relationships in CEM (https://w3id.org/CEMontology)

Ontology specification

- **Independent of crime legal representation** (that differs from country to country)
- Based on the **5W1H journalistic questions** related to objects (**`What`** has been stolen?, **`What`** weapon was used to commit the robbery?), participants (**`Who`** was involved as author or victim?), place and time (**`Where`** and **`When`** did the crime take place?), cause (**`Why`** did it happen?), procedure (**`How`** did it happen?)
- 28 different types of crimes modeled from the **International Classification of Crime for Statistical Purposes**
- Linked to the **news articles** reporting the events to enable comparison over different newspapers and media
- Compliant with the **FAIR** principles

Scan to visit CEM's documentation

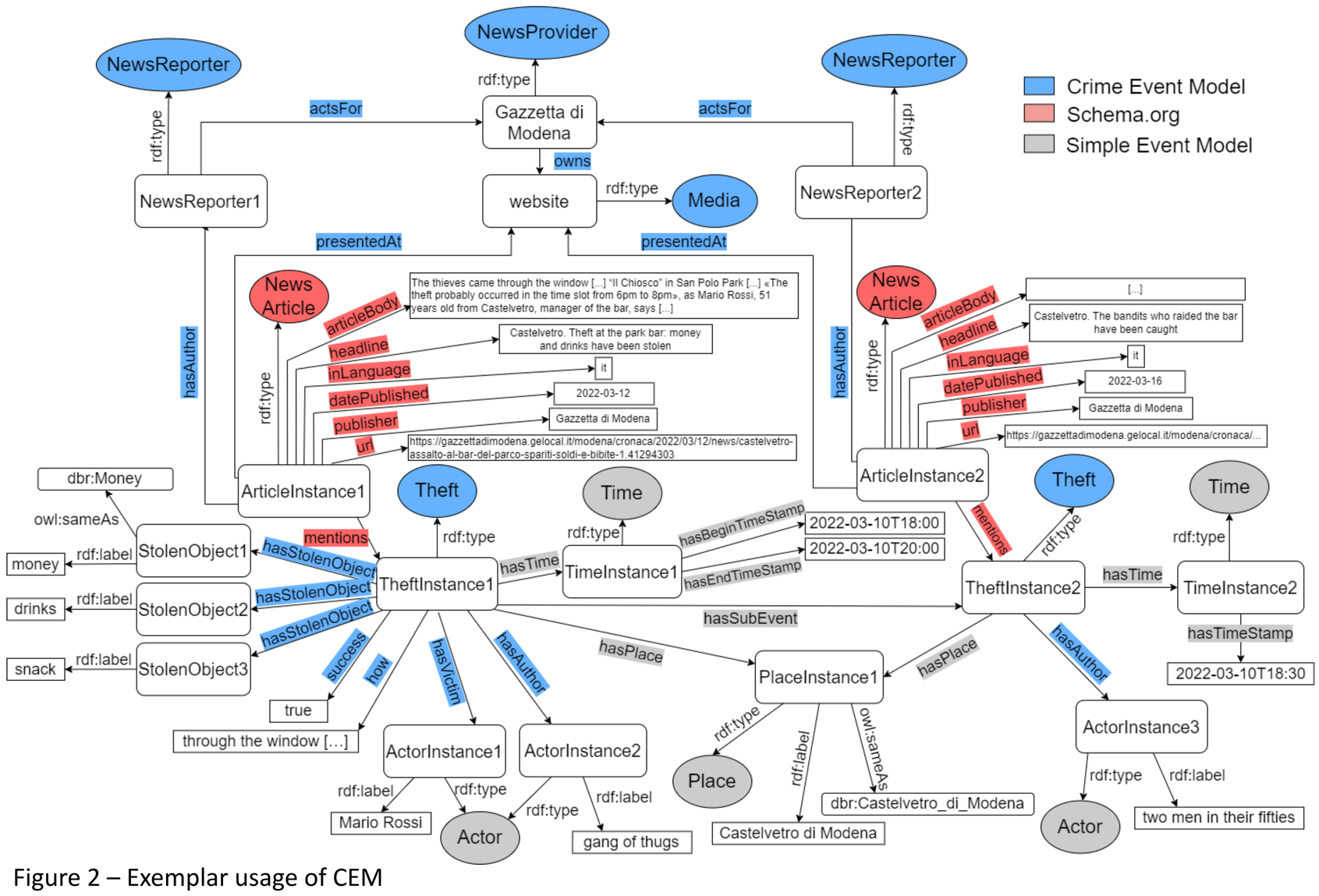

Figure 2 – Exemplar usage of CEM

Example of use

- **Italian Crime News dataset** (https://github.com/federicarollo/Italian-Crime-News)
  - 10,395 news articles
  - 13 types of crimes
  - 1 local newspaper: "Gazzetta di Modena"

Link to external semantic resources: instances can be connected to the concepts of semantic networks like *WordNet*, *ConceptNet* or *BabelNet*, *DBpedia*, *Wikidata* to enable further analysis (e.g., money thefts against women near supermarkets).

Scan to explore the dataset

Conclusion and Future Work

Extracting structured data from online sources and interconnecting them using CEM allow **events relationships extraction**, **trends identification** and **event recommendation**.

Further steps:

- **event extraction** on news articles for automatic population of CEM
- **community detection** and **pattern recognition** approaches for event-based analysis.

References

1. Rollo, F., Po, L. (2022). Knowledge Graphs for Community Detection in Textual Data. In: Villazón-Terrazas, B., Ortiz-Rodriguez, F., Tiwari, S., Sicilia, MA., Martín-Moncunill, D. (eds) Knowledge Graphs and Semantic Web. KGSWC 2022. Communications in Computer and Information Science, vol 1686. Springer, Cham. https://doi.org/10.1007/978-3-031-21422-6_15
2. Rollo, F., Bonisoli, G., Po, L. (2022). Supervised and Unsupervised Categorization of an Imbalanced Italian Crime News Dataset. In: Ziemba, E., Chmielarz, W. (eds) Information Technology for Management: Business and Social Issues. FedCSIS-AIST ISM 2021 2021. Lecture Notes in Business Information Processing, vol 442. Springer, Cham. https://doi.org/10.1007/978-3-030-98997-2_6
3. Rollo, Federica, and Po, Laura. (2020) Crime Event Localization and Deduplication. In: Pan J.Z. et al. (eds) The Semantic Web – ISWC 2020. ISWC 2020. Lecture Notes in Computer Science, vol 12507. Springer, Cham. https://doi.org/10.1007/978-3-030-62466-8_23