

GRACE: Graph-based Representation and Analysis for Crime Exploration

Gianpaolo Iuliano^{1,*†}, Grazia Margarella^{1,†}, Giuseppe Polese^{1,†} and Roberto Stanzione^{1,†}

¹University of Salerno, via Giovanni Paolo II, 132, Fisciano (SA), 84084, Italy

Abstract

Criminal investigations increasingly rely on the analysis of vast amounts of unstructured textual data, such as crime reports, social media content, and open-source intelligence. While recent advances in Natural Language Processing (NLP) have shown promise in automating parts of this process, existing approaches often lack explicit, interactive knowledge representations that investigators can directly manipulate and enrich. In this paper, we envision GRACE, a framework that combines automated information extraction, external data enrichment, and interactive visualization to support investigative workflows. Starting from unstructured crime reports, GRACE applies named entity recognition and relation extraction to construct a structured knowledge graph, which is further augmented with contextual attributes from external sources. Investigators can explore and refine this evolving graph through an interactive interface that supports natural language querying, dynamic visualization of connections, and manual correction or enrichment of knowledge. To show the potential of GRACE in supporting the investigation process, we provide a use case related to a fictional crime investigation. While still conceptual, our proposal aims to inspire the development of practical investigative tools that bridge automation and human expertise, ultimately enhancing the efficiency and effectiveness of criminal investigations.

Keywords

Criminal Investigations, Natural Language Processing, Information Extraction, Graphs, Interactive Visualization

1. Introduction

Criminal investigations require law enforcement agencies to collect, analyze, and interpret vast amounts of information. In recent years, the scale and complexity of this task have grown dramatically, as law enforcement agencies are confronted with millions of crime incidents every year. For instance, in the United States alone, over 14 million criminal offenses were reported in 2024¹. This huge volume of cases is compounded by the increasing set of digital traces, social media communications, and open-source intelligence that investigators must consider to fully contextualize evidence. To cope with this complexity, researchers and practitioners have increasingly explored computational methods to support the investigation process. For example, recent approaches have tackled problems such as determining the most influential members of a criminal group [1], preventing online crimes [2], or sorting and identifying relevant artifacts [3]. In parallel, large-scale European Union-funded projects such as TITANIUM² and TENSOR³ have also highlighted the growing interest in this direction.

At the heart of crime investigation processes lie crime reports, which serve as the primary documentation of incidents and provide investigators with the factual basis for understanding criminal activity. Each report is not only a narrative of a specific event but also a potential piece of a much larger puzzle that may reveal patterns, networks, or emerging threats. However, the sheer volume and

1st Workshop on supporting CRime reSolution Through Artificial Intelligence (CRISTAIN), held in conjunction with CHITALY 2025, Salerno, Italy, October 6–10

*Corresponding author.

†These authors contributed equally.

✉ giaiuliano@unisa.it (G. Iuliano); gmargarella@unisa.it (G. Margarella); gpolese@unisa.it (G. Polese); rstanzione@unisa.it (R. Stanzione)

ORCID: 0000-0001-5049-0317 (G. Iuliano); 0009-0007-5520-2872 (G. Margarella); 0000-0002-8496-2658 (G. Polese); 0000-0001-9212-821X (R. Stanzione)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://www.fbi.gov/news/press-releases/fbi-releases-2024-reported-crimes-in-the-nation-statistics>

²<https://titanium-project.eu/>

³<https://tensor-project.eu/>

heterogeneity of these documents make the extraction of relevant information and the identification of relationships a slow and demanding task. Recent advances in Natural Language Processing (NLP) represent a major opportunity to address this challenge, enabling to automatically extract knowledge from large volumes of unstructured text and uncover hidden relationships. Recent works have begun to explore the use of NLP to support investigative processes [4, 5]. However, existing systems often lack an explicit representation of knowledge that can be directly manipulated by investigators. Without such representations, it becomes difficult for investigators to explore the data beyond the system’s predefined outputs. Furthermore, current tools offer limited interactivity: investigators may receive extracted results or predictions, but they are not empowered to query the data in natural language, to visualize connections dynamically, or to correct and enrich the knowledge base with their own insights. Finally, the integration of external contextual knowledge may further support and ease the investigations.

Motivated by these gaps, in this paper we envision an investigative support tool, namely GRACE (*Graph-based Representation and Analysis for Crime Exploration*), that combines automated information extraction, data enrichment, and interactive visualization. Starting from textual crime reports, GRACE applies Named Entity Recognition (NER) and Relation Extraction (RE) to identify relevant entities and their relationships. These elements are represented as nodes and relationships in a graph, creating an initial structured representation of the case. This graph is then enriched through the integration of external knowledge sources, transforming it into an enriched graph where nodes and edges are associated with attributes, and can be dynamically updated when new reports are available. Moreover, GRACE provides an interface that enables investigators to explore, query, and refine the knowledge base. In addition, it supports natural language querying, allowing investigators to easily formulate questions. The visual environment also allows dynamic exploration of connections and the possibility to add or correct information. To summarize, the key novelties that characterize GRACE are the following:

- It allows to automatically enrich the graph obtained from the reports with external knowledge;
- It allows to dynamically update the aforementioned graph when new case reports are available;
- It provides users with an interface that they can analyze and query using natural language and/or predefined queries.

The rest of this paper is organized as follows: Section 2 discusses related works, while Section 3 provides an overview of the problem. Section 4 details GRACE and its components, while Section 5 provides a use case to show its potential. Finally, Section 6 provides closing remarks and future directions.

2. Related Work

In recent years, several research efforts have proposed computational approaches to support forensic reasoning and crime investigation. In [1], the authors introduce CISRI, a graph-based framework aimed at analyzing the structural properties of criminal networks, enabling the identification of important actors and relational patterns. On the other hand, in [2], the authors propose an approach that, starting from unstructured social media content, builds a knowledge graph, thereby providing a structure that captures relationships and that can be used to uncover hidden patterns and relationships. Graphs are also leveraged in [6], in which the authors focus on legal reasoning about evidences. In particular, they explore the use of support graphs as an intermediate representation to translate Bayesian networks into argumentative structures, thereby improving the interpretability of probabilistic evidence while preserving the underlying independence information. Finally, another graph-based approach was proposed in [7], in which the authors propose a novel graph model for network forensics, where evidence graphs support both local and global reasoning about attack scenarios and enable interactive hypothesis testing to uncover implicit attacker behavior. Recently, the research community explored the use of NLP techniques in their approaches. In [4], the authors present an early warning system based on NLP and knowledge discovery that assists law enforcement agencies in collecting and analyzing large volumes of unstructured textual data from sources such as the dark web. In particular, the authors leverage NER to build a Knowledge Repository, and then they provide a set of Knowledge Discovery

tools to perform tasks such as association rule mining and clustering. On the other hand, in [5] the authors combine NLP and graph-based methods to extract and structure relations from crime reports on violence against women in India. Named entities are identified via NER, and relationships are captured through hierarchical graph clustering.

The works above illustrate the potential of both graph-based and NLP-driven approaches in supporting forensic reasoning and criminal investigations. However, only a limited number of studies attempt to combine the two perspectives, leaving open opportunities for designing more refined systems. Moreover, none of the reviewed works addresses the design of an interactive interface to enable investigators to actively manipulate and act upon the extracted data and knowledge. Finally, none of the existing approaches considers the possibility of dynamically updating the graph, either manually or automatically. This aspect is crucial in real investigative scenarios, where new reports continuously emerge and the knowledge base must evolve accordingly. With GRACE, we envision a more complete framework that includes all these features, thereby providing a valuable support to crime investigations.

3. Problem overview

In the investigative context, textual reports are one of the main sources of information, as they contain detailed descriptions of crime scenes, testimonies, and observations. These documents contain a wide variety of entities and relationships, such as people, places, events, and objects, expressed in natural language. The analysis of such texts is crucial for detectives, who must extract relevant elements, relate them to each other, and construct a coherent representation of events. However, this process remains largely manual, time-consuming, and cognitively demanding. Despite the significant effort required for this process, several studies in the field of criminology have shown that the structured extraction of entities and relationships from investigative texts can drastically reduce the risk of information loss and facilitate the identification of recurring patterns [8, 9].

Crime reports often lack a formal representation that allows immediate correlations between documents, resulting in fragmented information and making it difficult to identify indirect connections between people and places, or between events separated in time. Furthermore, the absence of integration with external knowledge bases, such as DBpedia [10] and Wikidata [11], limits the possibility of contextualizing and enriching data, for example, by disambiguating entities with common names or linking an organization mentioned in reports to public domain information. Another critical issue with the tools currently available concerns the difficulties experienced by non-expert users in using them [12]. Although these technologies have demonstrated great potential in various application domains, their use often remains confined to highly specialized academic or industrial environments. Access to data is mediated by formal query languages, such as SPARQL[13], which require advanced technical skills and in-depth knowledge of the underlying models. Recent studies emphasize that modeling and interacting with graphs are complex activities that are not easily accessible to non-expert users, both because of the difficulty of mastering the formalisms and the absence of truly intuitive interfaces [14]. This technological gap translates into a significant barrier in the investigative context: detectives, despite being experts in reading and interpreting reports, generally do not have simple and immediate tools to transform such data into structured and navigable knowledge. This results in the need for solutions that reduce technical complexity, favoring natural and user-centered modes of interaction.

To address these limitations, GRACE aims to transform investigative reports into structured and navigable knowledge. It applies NER and RE to identify and connect entities such as people, places, and events, organizing them into a dynamic knowledge graph enriched with external resources, like Wikidata and DBpedia, enabling disambiguation and contextualization. GRACE is designed for detectives and non-technical users, providing an intuitive interface that allows them to explore, query, and modify the graph by adding, removing, or editing entities and relationships. Predefined queries support common investigative tasks, while a text-to-query component allows users to formulate natural language questions without expertise in formal query languages. This combination of automated extraction, semantic enrichment, interactive editing, and flexible querying, allows to reduce analysis time, lowers

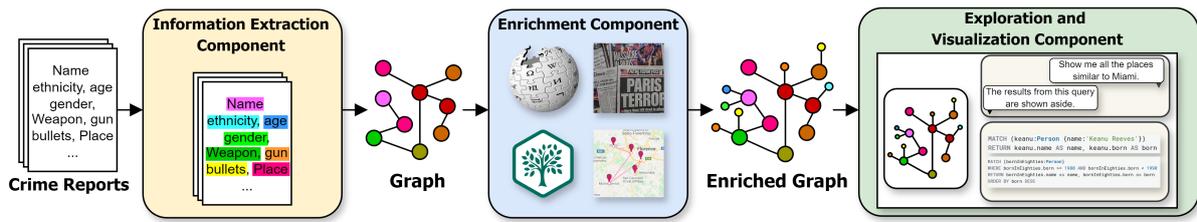


Figure 1: Overview of GRACE

cognitive load, and enhances the ability to detect connections and patterns.

4. The GRACE system

GRACE is structured as a modular pipeline that integrates NLP techniques, knowledge graph construction, and interactive visualization. Its main goal is to transform unstructured investigative reports into a structured, navigable representation that supports investigators in their analytical activities. As can be seen in Figure 1, GRACE consists of several components that allow the execution of the extraction and enrichment tasks, as well as the exploration of data. Initially, the investigative reports are fed into the *Information Extraction Component*, which applies NLP techniques to identify and extract data from the available documents. Once extracted, this information is organized into a graph, which allows it to be represented in a structured and easily navigable manner. This graph is then enriched through the *Enrichment Component*. Specifically, this component integrates external knowledge from resources such as Wikidata or DBpedia, enabling disambiguation and semantic enhancement of entities. This process increases the knowledge present in the graph, providing additional details and characteristics related to the various entities extracted. Once the graph has been constructed and enriched, it becomes the input for the *Exploration and Visualization Component*, which makes the graph accessible through an interactive graphical interface designed specifically for non-technical users. This interface allows investigators to explore the graph, manually modify its content, and execute queries.

In what follows, we detail GRACE’s main components, describing the techniques used in each of them.

4.1. Information Extraction Component

The *Information Extraction Component* deals with all text analysis processes aimed at identifying entities and relationships between them within investigative reports. In particular, this component takes text reports as input and applies NER and RE techniques to identify main entities, such as people, places, weapons, or events, as well as the relationships between them. To extract this information, the Information Extraction component can leverage fine-tuned Large Language Models (e.g., LLaMA [15], GPT [16]) or Machine Learning approaches designed for the criminological domain [17]. These approaches identify entities, such as people, criminal organizations, places, types of crime, and material evidence, and relationships between them, for example, linking a suspect to a specific crime or a criminal event to a specific location. The process involves a text pre-processing phase, followed by zero-shot or few-shot model inference, and subsequently post-processed aimed at normalizing entities and reducing false positives. At the end of this process, the user can manually interact with the results by adding or removing entities and relationships not extracted by them. The results are then organized into a graph, which allows for the representation of the case in a structured manner, facilitating the identification of patterns and meaningful connections.

4.2. Enrichment Component

The graph generated by the *Information Extraction Component* is provided as input for the second component of GRACE, i.e., the *Enrichment component*. This component extends the graph beyond the

information explicitly extracted from reports, resolving ambiguities and adding contextual knowledge. Its operation relies on entity linking techniques, which map textual mentions of people, places, or organizations to identifiers in external resources such as Wikidata or DBpedia. This process combines string similarity measures (e.g., Levenshtein distance), contextual embeddings derived from language models, and disambiguation rules based on co-occurring entities in the same document [18]. Once an entity is matched, the system uses SPARQL queries to retrieve relevant attributes and relations, such as alternative names, geographical coordinates, affiliations, or classifications. These are integrated into the graph as additional node properties or new edges, while maintaining metadata that distinguishes extracted information from externally enriched knowledge and the external knowledge bases.

In this way, this component transforms the graph from an exact representation of the reports into a semantically enhanced structure, where entities are enriched and connected through meaningful relationships extracted from both the reports and external knowledge bases.

4.3. Exploration and Visualization Component

The enriched graph produced by the previous component serves as the input for the Exploration and Visualization component. This component provides investigators with a versatile interface to explore, analyze, and manipulate the graph in a way that reflects both the original reports and the additional contextual information integrated during the enrichment. In particular, queries can be executed either using predefined templates, specifically designed to address common investigative scenarios, or through a text-to-query conversion module [19], which allows for the translation of natural language questions into formal queries without requiring technical expertise. Beyond querying, the interface enables investigators to actively manage the graph: they can add new nodes and edges as new information emerges, remove outdated or incorrect entries, and modify existing connections to ensure that the representation remains accurate and up-to-date. By combining flexible querying with direct graph manipulation, this interactive layer not only facilitates the rapid retrieval of relevant information but also supports iterative analysis, hypothesis testing, and the identification of previously unrecognized connections, making the tool a comprehensive support system for investigative reasoning.

5. Use Case

In order to illustrate the features of the envisioned tool, we present a use case focused on analyzing a fictional police report. This use case demonstrates how the tool GRACE would define an enriched graph for further knowledge description. The workflow delineated below incorporates automated information extraction, graph construction, a proposed query interface, and graph editing. To illustrate this, we considered the generation of the graph based on a fictional report of the American serial killer Jeffrey Dahmer. The initial text constitutes a police report that documents the apprehension of the suspect after the escape of a victim. Thereby revealing a scene characterized by the presence of a substantial amount of evidence, including human remains, chemical substances, and photographic documentation.

After the selection of a new report, the *Information Extraction Component* is the first to be applied to it. The application of NER results in the extraction of several entities, including *Persons*, *Locations*, and *Evidences*. As shown in Figure 2, upon the selection of the report, the analysis results are both highlighted in the text and grouped by entity categories. For instance, a person could be exemplified by Jeffrey Dahmer. The suspect is related to the victim, Tracy E., who reported that he had been threatened with a knife. So this situation is described by the entities *Jeffrey Dahmer*, *Tracy E.*, and *Knife*, which are connected through the relationships “*reported_against*” and “*threatened_with*”. In the sample graph are also described the different evidences and the requested forensic analysis. In addition to this, the user can also interact to remove or highlight some of the missed entities, and later confirm the validity of the results for the graph definition.

The generated graph is then enriched through the *Enrichment Component*, which is callable by the database icon in the top-left corner of the interface at any time. As illustrated in Figure 3, the user can see which sources are analyzed, including, but not limited to, other case reports, ontologies, newspapers,

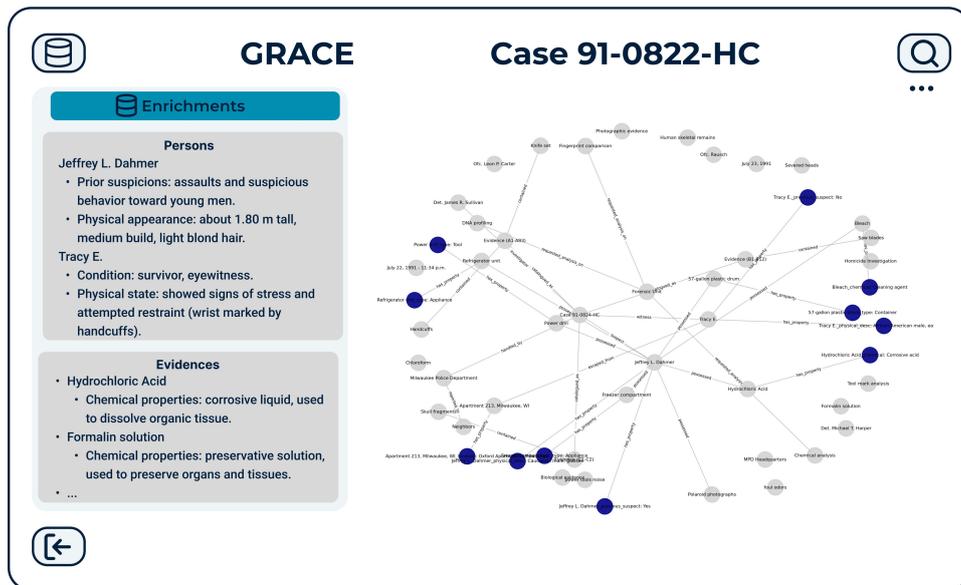


Figure 4: Mock-up for the resultant graph in the GRACE Enrichment Component.

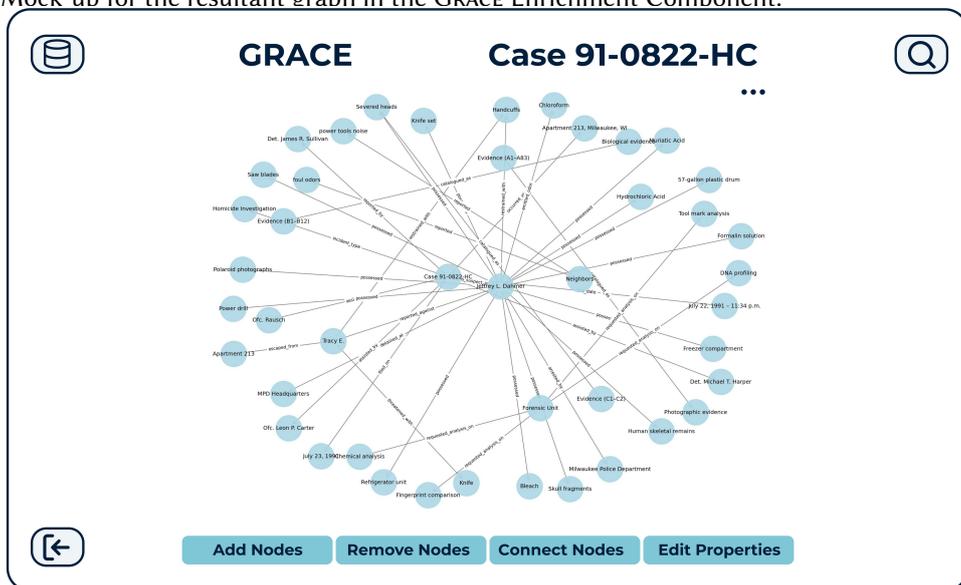


Figure 5: Mock-up for graph interaction in GRACE Exploration and Visualization Component.

information provided by different sources. In addition, the graph representation with the proposed querying features can support the investigators in discovering insights and uncovering hidden relations.

Furthermore, GRACE also facilitates the concurrent analysis of multiple reports in the Information Extraction Component, ensuring the consistency of the extracted entities, which are subsequently represented within the graph. For instance, in such a scenario, the case of Jeffrey Dahmer could also be supported by multiple reports from the victims' discoveries. As a matter of fact, during the exploration of the graph, the user can select standard queries, such as the first in Figure 6, which allows to extract the "Cases where Jeffrey Dahmer is a suspect" and enables the reconstruction of the report identifiers to visualize which entities are involved in which case. On the other hand, the Text2Query feature could deal with a prompt such as "Consider all incidents involving Jeffrey Dahmer and the chemical substances discovered in his apartment, return the associated victims and the common traits between them", which has to interact with the several victims' information in different case reports. These different usage modes of GRACE demonstrate its versatility in different investigation contexts, either those involving a single report or multiple ones.

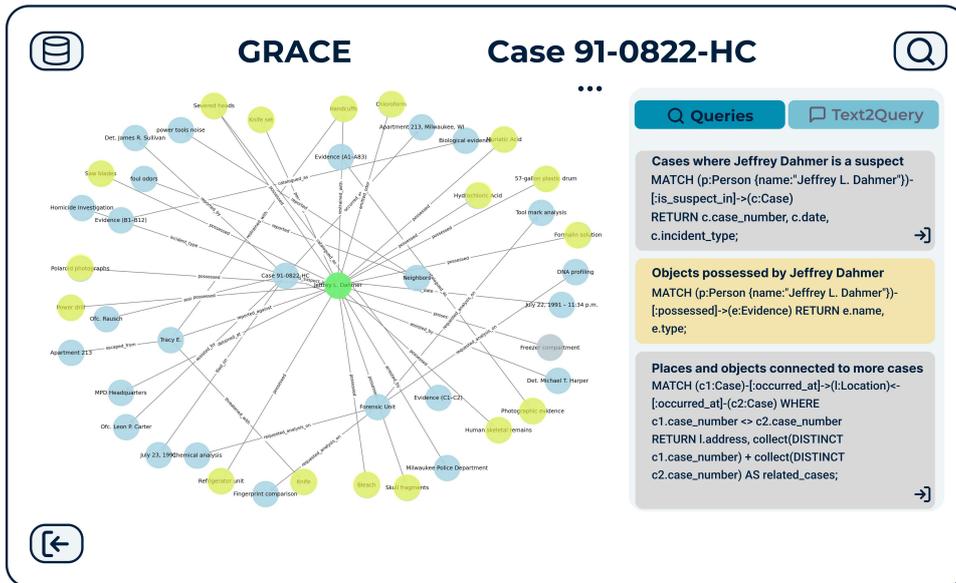


Figure 6: Mock-up for standard queries in the GRACE Exploration and Visualization Component.

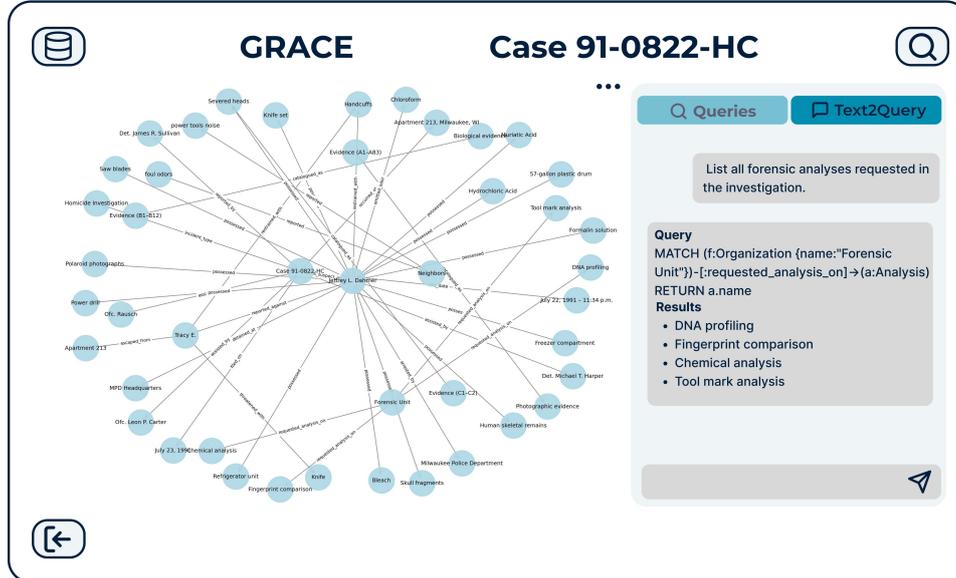


Figure 7: Mock-up for Text2Query Interaction of the GRACE Exploration and Visualization Component.

6. Conclusions

In this paper, we envision GRACE, a framework that integrates automated information extraction, external data enrichment, and interactive visualization to support criminal investigations. GRACE empowers investigators not only to visualize and explore complex connections among entities, but also to manipulate, enrich, and query the knowledge base, bridging the gap between automation and human expertise. We also provided a fictional use case to show the potential impact of GRACE on criminal investigations. In the future, we aim to evaluate a prototype of GRACE to evaluate it within real-world scenarios to assess its usability and its impact on investigative workflows. In addition, integration with multimedia data could further expand the scope of analysis beyond textual reports. Finally, future developments should also ensure compliance with privacy requirements and legal regulations, so that GRACE can be deployed responsibly within law enforcement contexts.

Acknowledgments

This study was funded by Ministero dell'Università e della Ricerca (MUR) of Italy in the context of the project denoted as BLOODSTAIN in the program PRIN 2022 (grant number D53D23008660006).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] M. Alzaabi, K. Taha, T. A. Martin, Cisri: A crime investigation system using the relative importance of information spreaders in networks depicting criminals communications, *IEEE Transactions on Information Forensics and Security* 10 (2015) 2196–2211.
- [2] O. Elezaj, S. Y. Yayilgan, E. Kalemi, L. Wendelberg, M. Abomhara, J. Ahmed, Towards designing a knowledge graph-based framework for investigating and preventing crime on online social networks, in: *International Conference on e-Democracy*, Springer, 2019, pp. 181–195.
- [3] M. Al Fahdi, N. Clarke, S. Furnell, Towards an automated forensic examiner (afe) based upon criminal profiling & artificial intelligence (2013).
- [4] C. Fernandez-Basso, K. Gutiérrez-Batista, J. Gómez-Romero, M. D. Ruiz, M. J. Martin-Bautista, An ai knowledge-based system for police assistance in crime investigation, *Expert Systems* 42 (2025) e13524.
- [5] P. Das, A. K. Das, J. Nayak, D. Pelusi, W. Ding, A graph based clustering approach for relation extraction from crime data, *IEEE Access* 7 (2019) 101269–101282.
- [6] S. T. Timmer, J.-J. C. Meyer, H. Prakken, S. Renooij, B. Verheij, Explaining legal bayesian networks using support graphs, in: *Legal Knowledge and Information Systems*, IOS Press, 2015, pp. 121–130.
- [7] W. Wang, T. E. Daniels, A graph based approach toward network forensics analysis, *ACM Transactions on Information and System Security (TISSEC)* 12 (2008) 1–33.
- [8] J. W. Johnsen, K. Franke, The impact of preprocessing in natural language for open source intelligence and criminal investigation, in: *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 2019, pp. 4248–4254.
- [9] M. Kaur, M. Saini, Role of artificial intelligence in the crime prediction and pattern analysis studies published over the last decade: a scientometric analysis, *Artificial Intelligence Review* 57 (2024) 202.
- [10] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, Dbpedia: A nucleus for a web of open data, in: *international semantic web conference*, Springer, 2007, pp. 722–735.
- [11] D. Vrandečić, Wikidata: A new platform for collaborative data collection, in: *Proceedings of the 21st international conference on world wide web*, 2012, pp. 1063–1064.
- [12] E. Kuric, J. D. Fernández, O. Drozd, Knowledge graph exploration: a usability evaluation of query builders for laypeople, in: *International Conference on Semantic Systems*, Springer, 2019, pp. 326–342.
- [13] J. Pérez, M. Arenas, C. Gutierrez, Semantics and complexity of sparql, *ACM Transactions on Database Systems (TODS)* 34 (2009) 1–45.
- [14] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, *ACM Computing Surveys (Csur)* 54 (2021) 1–37.
- [15] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, et al., The llama 3 herd of models, *arXiv preprint arXiv:2407.21783* (2024).
- [16] R. Islam, O. M. Moushi, Gpt-4o: The cutting-edge advancement in multimodal llm, in: *Intelligent Computing-Proceedings of the Computing Conference*, Springer, 2025, pp. 47–60.
- [17] P. Das, A. K. Das, A two-stage approach of named-entity recognition for crime analysis, in: *2017 8th international conference on computing, communication and networking technologies (ICCCNT)*, IEEE, 2017, pp. 1–5.
- [18] I. Guellil, A. Garcia-Dominguez, P. R. Lewis, S. Hussain, G. Smith, Entity linking for english and other languages: a survey, *Knowledge and Information Systems* 66 (2024) 3773–3824.
- [19] I. Mandilara, C. M. Androna, E. Fotopoulou, A. Zafeiropoulos, S. Papavassiliou, Decoding the mystery: How can llms turn text into cypher in complex knowledge graphs?, *IEEE Access* (2025).