Constructing Cybersecurity Knowledge Graphs for Hybrid LLM–Graph Reasoning on Vulnerabilities

Julio Vizcarra*, Yuta Gempei, Yanan Wang, Takamasa Isohara and Mori Kurokawa

KDDI Research, Inc., Japan

Abstract

In cybersecurity, the threat landscape is composed of complex relations among security data and constantly evolves. To address this challenge, this paper presents a framework for constructing and reasoning over cybersecurity knowledge graphs (KGs) derived from vulnerability reports. Our approach analyzes textual content and structured data sources. To enhance causal reasoning, we explicitly model key causal factors as structured entities and relationships. The resulting KG is further enriched through augmentation using DBpedia, integrating external knowledge to enhance connectivity and context. We evaluate the impact of this augmentation through a comparison, contrasting the content of the original and the augmented graphs. Experimental results demonstrate that the Graph-LLM approach, with augmentation, enhances link prediction and produces higher-quality Question Answering (QA) compared to using report descriptions alone. We demonstrate a hybrid reasoning setup integrating LLM-based language understanding with graph inference to answer cybersecurity queries.

Keywords

Knowledge graph, text mining, causality, LLM, cybersecurity vulnerabilities, DBpedia.

1. Introduction

Cybersecurity is a specialized domain where analysts must interpret vast, evolving data to detect threats, understand vulnerabilities, and respond appropriately [1]. Knowledge graphs (KGs) have emerged as an effective tool for modeling this structured information, providing a semantic foundation for threat analysis, attack attribution, and decision-making [2]. Despite their advantages, constructing and maintaining cybersecurity knowledge graphs is resource-intensive, and automated methods are still in development [3]. Large language models (LLMs) excel at natural language understanding and entity extraction but struggle with precise symbolic reasoning and complex graph operations like multi-hop inference and causal chaining [4]. In contrast, graph-based reasoning is highly effective at traversing causal and semantic relationships, which are critical for cybersecurity tasks such as attack path analysis and threat correlation [5]. The combination of LLMs' natural language understanding with structured graph reasoning can offer a powerful hybrid solution, leveraging the strengths of both paradigms [6].

1.1. Our Approach

To address the challenges outlined above, we propose a hybrid framework that combines strong language understanding of LLMs with the semantic expressiveness of knowledge graphs for cybersecurity threat understanding. To this end, the KG is first constructed by analyzing structured and unstructured data to discover associations in vulnerability reports CPE [7], CVE [8], and CWE [9], where CPE identifies affected products and versions, CVE denotes specific vulnerabilities in software and hardware, and CWE classifies the underlying weakness types. We also introduce a causal modeling schema that explicitly represents causal elements, enhancing the ability of the graph to support cause-effect reasoning. To improve coverage and semantic extension, we augment the KG using DBpedia [10], enriching it with background knowledge (DBpedia resources) and semantic relations (ontological terms). This augmentation improves connectivity and enables higher-quality inference. Experiments show the

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan

*Corresponding author.

xju-vizcarra@kddi.com (J. Vizcarra); yu-genpei@kddi.com (Y. Gempei); wa-yanan@kddi.com (Y. Wang); ta-isohara@kddi.com (T. Isohara); mo-kurokawa@kddi.com (M. Kurokawa)

augmented Graph-LLM improves link prediction and Question Answering (QA) quality over report descriptions alone. Finally, we present a hybrid pipeline that integrates LLM outputs with KG inference to answer cybersecurity queries, thereby enhancing reasoning through knowledge graph retrieval.

2. Methodology

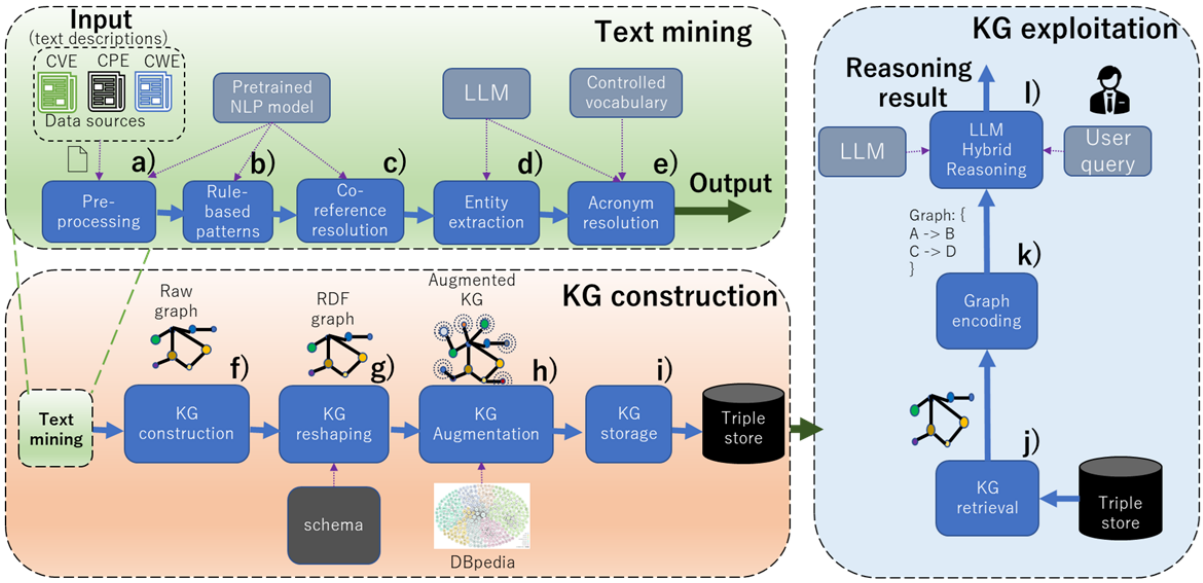

Figure 1: Pipelines that encompass the KG creation and exploitation. 1) Text mining extracts valuable information from text descriptions. 3) KG creation transforms the results of text mining and available structured data into a KG, represented as an RDF graph. 3) KG exploitation enables LLM-graph reasoning in response to user queries.

The methodology depicted in Fig. 1 comprises several steps grouped into pipelines: 1) text mining extracts content from the text descriptions, 2) KG creation builds the KG combining structured data and content from the text mining, and 3) KG exploitation performs reasoning on the KG created. For LLM processing, we used LLaMA 3.3 70B instruct model [11].

1. Text mining.
   a) In pre-processing, vulnerability reports are retrieved, parsed, and cleaned.
   b) Rule patterns and the syntactic tree. This process is used to identify concepts and relationships. It is performed as a fallback option of LLM entity extraction.
   c) Co-reference resolution is computed to extract the reference for the same entity in the text.
   d) Entity extraction. This process is handled by the LLM, which identifies causal factors: entities, relationships, states, actors, objects, and properties.
   e) Acronym resolution is performed using specialized dictionaries and LLM knowledge.
2. The KG construction.
   f) The KG construction organizes the data collected through text mining into a raw graph (i.e., a graph without Resource Description Framework (RDF) notation [12]).
   g) The KG reshaping converts the raw graph into an RDF graph [13], which is a foundational data model for representing knowledge as triples. The KG is based on a schema (ontology).
   h) The graph is augmented using concepts and relations from DBpedia. The LLM selects the most appropriate concept during expansion to create coherent graph excerpts.
   i) The graph created is stored in a Virtuoso open-source triple store [14].

3. KG exploitation.
    j) In the KG retrieval step. SPARQL queries [15] retrieve content from the KG stored in the triple store based on the user request using multi-hop expansion.
    k) The graphs retrieved are encoded into text for the LLM's processing.
    l) The LLM hybrid reasoning combines graph data and the user query to perform reasoning. .

3. Experimental results

3.1. Knowledge Graph Creation and Augmentation

An example of knowledge graph creation and augmentation is presented. A graph excerpt derived from structured information and its augmentation is presented in Fig. 2. The main elements of the graph created are highlighted.

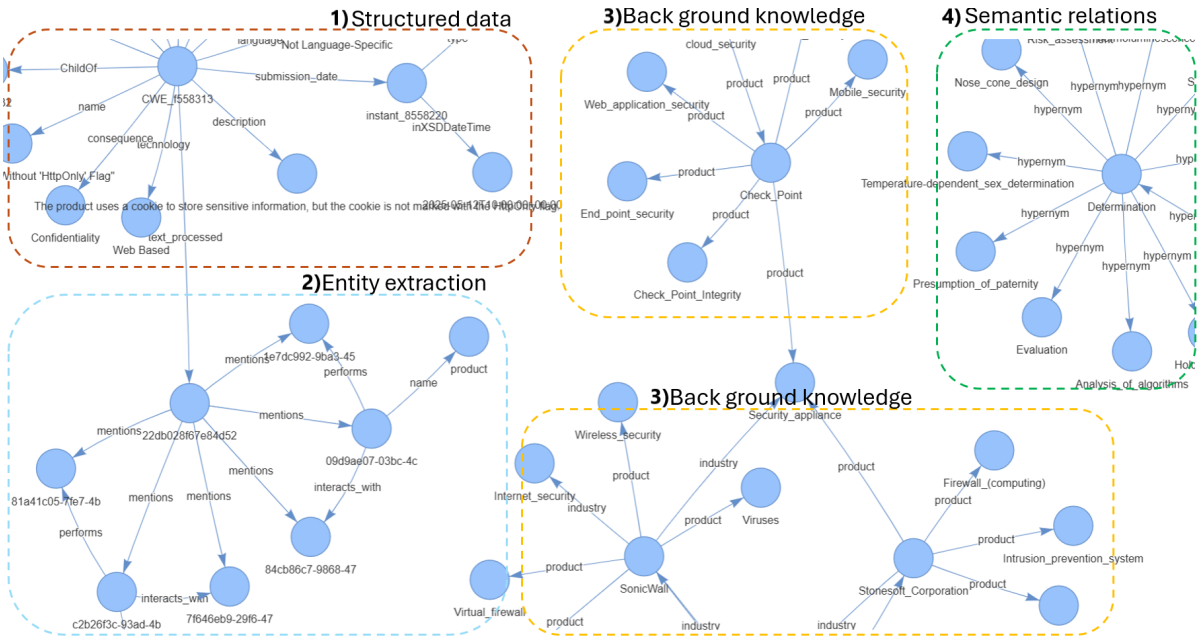

Figure 2: Final KG excerpt for the CWE report presents content from 1) structured data, 2) causal factors (actor, action, object, properties, states) from entity extraction, 3) background knowledge from DBpedia (similar products and industries), and 4) cybersecurity terminology and semantic relations.

3.2. LLM-Graph Reasoning and Exploitation.

A response generated by LLM–graph reasoning is shown in Table 1. In this hybrid approach, the LLM analyzes the user query alongside the graph content retrieved based on terms from the question. The first row presents the result of the LLM reasoning using only plain text (i.e., the report description). In contrast, the graph-LLM response includes versions connected to this report.

Additionally, the visualization of the KG exploitation is depicted in Fig.3. Once the query is entered, the user can select a specific vulnerability report (CWE, CVE, or CPE) to perform the reasoning. The user can define the number of hops to expand the query and choose whether the expansion includes conceptual relations or only directly related reports.

3.3. Evaluation

We measured new relation discovery after KG augmentation using link prediction with AmpliGraph [16]. On the CWE subgraph, TransE, ComplEx, and DistMult [13] were evaluated; ComplEx performed

Table 1
Example query and reasoning results. The query is answered: a) by the LLM using only the plain text from the report descriptions, and b) by the graph-LLM reasoning process using graph data. The graph-LLM option provides more relevant content by leveraging information from the knowledge graph.

Query	*Reasoning result*
Where is the issue located in the affected Extreme Networks EXOS versions	**Ground truth answer**: In the Web GUI. **Graph-LLM answer**: Based on the graph data, the issue is located in the Web GUI of Extreme Networks EXOS versions before v.22.7 and before v.30.2. The problem stems from the failure of the Web GUI to restrict URL access, allowing attackers to access sensitive information or escalate privileges.

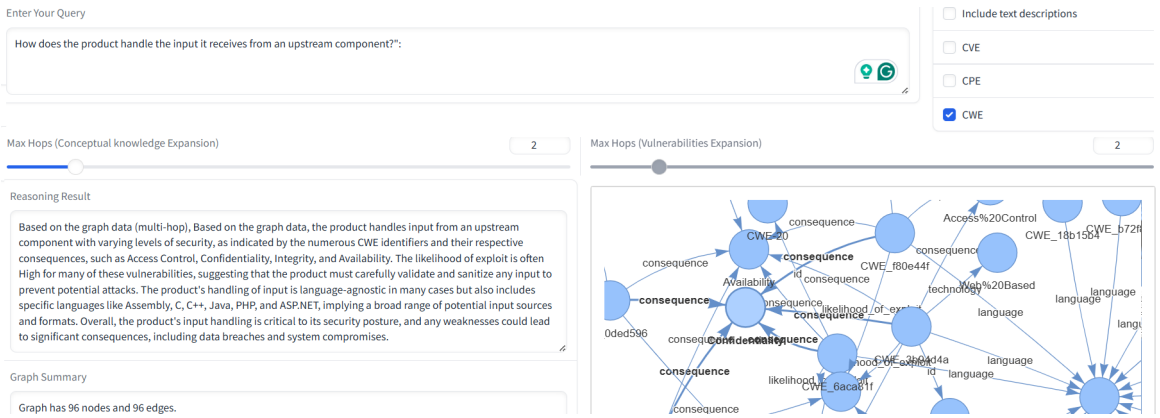

Figure 3: The UI supports Graph-LLM reasoning to enhance explainability. Users can use the interface to validate the KG retrieved and used by the LLM. It is also possible to adjust the number of hops for QA expansion (conceptual or based on similar connected reports) and to specify particular security reports.

best, and the KG with text mining content (no DBpedia) achieved the highest results (Table 2). The augmented KG showed low performance because the LLM struggled with KG excerpt selection; that is, it chose an incorrect meaning for the expanded concept on DBpedia, adding noise and negatively affecting link prediction. Using a different LLM that handles this task better might improve performance.

Table 2
In link prediction task, KG augmented with content produced by text mining content improved the results compared to the baseline (structured). Mean Rank lower is better, and the other metrics higher are better.

Link prediction using ComplEx (best model)	*Mean Rank*	*MRR*	*Hits@1*	*Hits@10*
CWE Structured (base line)	426.32	0.2493	0.1815	0.3705
CWE Augmented No DBpedia (ours)	**395.06**	**0.3015**	**0.2330**	**0.4310**
CWE Augmented Full (ours)	468.08	0.2808	0.2075	0.4255

In addition, the quality of the responses was measured by Gemma 3 27B model [17] with a multi-hop expansion of 5 levels. LLM with ground truth refers to answers derived from the text descriptions in vulnerability reports. In contrast, Graph LLM responses are generated using hybrid reasoning. In the evaluation, Graph LLM produced showed the best performance (Table 3).

Table 3
Graph-LLM shows improved performance over LLM with ground truth security reports using 1000 CWE entries and 3000 Question-Answer pairs.

Category	Dataset	Accuracy	Relevance	Clarity	Completeness
CWE	LLM with ground truth	0.648	0.520	0.530	0.266
	Graph-LLM (ours)	**0.802**	**0.688**	**0.670**	**0.630**

4. Conclusions and Future Work

This poster presents a hybrid framework for cybersecurity threat reasoning that combines automatically constructed knowledge graphs with the reasoning capabilities of large language models (LLMs). The framework created graphs that represent causal factors and was enriched through DBpedia augmentation. To overcome LLM limitations in complex graph operations, we developed a dedicated subgraph extraction process. This module lets LLMs handle language comprehension and logical reasoning, while the graph module performs structured inference. Together, this hybrid pipeline provides a scalable and interpretable solution for addressing complex cybersecurity queries, with potential applications extending beyond this domain. As a practical application, the proposed visualization system provides a tool for QA analysis. As part of our future work, we plan to evaluate additional queries and scenarios.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] M. Syafrizal, S. R. Selamat, N. A. Zakaria, Analysis of cybersecurity standard and framework components, International Journal of Communication Networks and Information Security 12 (2020) 417–432.

[2] L. F. Sikos, Cybersecurity knowledge graphs, Knowledge and Information Systems 65 (2023) 3511–3531.

[3] M. Hofer, D. Obraczka, A. Saeedi, H. Köpcke, E. Rahm, Construction of knowledge graphs: Current state and challenges, Information 15 (2024) 509.

[4] A. Patil, Advancing reasoning in large language models: Promising methods and approaches, arXiv preprint arXiv:2502.03671 (2025).

[5] C. Zhong, J. Yen, P. Liu, R. F. Erbacher, Automate cybersecurity data triage by leveraging human analysts' cognitive process, in: 2016 IEEE 2nd International Conference on big data security on cloud (BigDataSecurity), IEEE International Conference on high performance and smart computing (HPSC), and IEEE International Conference on intelligent data and security (IDS), IEEE, 2016, pp. 357–363.

[6] S. Pan, Y. Zheng, Y. Liu, Integrating graphs with large language models: Methods and prospects, IEEE Intelligent Systems 39 (2024) 64–68.

[7] National Institute of Standards and Technology, CPE: Common Platform Enumeration, https://nvd.nist.gov/products/cpe, 2024. Accessed: 2025-05-15.

[8] The MITRE Corporation, CVE: Common Vulnerabilities and Exposures, https://cve.mitre.org, 2024. Accessed: 2025-05-15.

[9] The MITRE Corporation, CWE: Common Weakness Enumeration, https://cwe.mitre.org, 2024. Accessed: 2025-05-15.

[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] AI@Meta, Llama 3 model card (2024). URL: https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.
[12] S. Decker, S. Melnik, F. Van Harmelen, D. Fensel, M. Klein, J. Broekstra, M. Erdmann, I. Horrocks, The semantic web: The roles of xml and rdf, IEEE Internet computing 4 (2000) 63–73.
[13] V. Bonstrom, A. Hinze, H. Schweppe, Storing rdf as a graph, in: 2003 First Latin American Web Congress, IEEE, 2003, pp. 27–36.
[14] OpenLink Software, Virtuoso Open-Source Edition, https://virtuoso.openlinksw.com/, 2024. Accessed: 2025-05-01.
[15] B. DuCharme, Learning SPARQL: querying and updating with SPARQL 1.1, " O'Reilly Media, Inc.", 2013.
[16] L. Costabello, A. Bernardi, A. Janik, A. Creo, S. Pai, C. L. Van, R. McGrath, N. McCarthy, P. Tabacof, AmpliGraph: a Library for Representation Learning on Knowledge Graphs, 2019. URL: https://doi.org/10.5281/zenodo.2595043. doi:10.5281/zenodo.2595043.
[17] G. Team, Gemma 3 (2025). URL: https://goo.gle/Gemma3Report.