**Mykola Stasiuk, Bohdan Pavlyshenko***
[1] Ivan Franko National University of Lviv, Lviv, Ukraine. ORCID: 0009-0006-5256-2103, mykola.stasiuk@lnu.edu.ua
[2] Ivan Franko National University of Lviv, Lviv, Ukraine. ORCID: 0000-0001-9515-3488, bohdan.pavlyshenko@lnu.edu.ua

# Visualizing and Validating Multilingual Thematic Structures in Knowledge Graphs

***Abstract.*** *This study presents an extended methodological framework for exploring and validating thematic structures within multilingual cybersecurity discourse. Building upon bilingual knowledge graphs that integrate entities and relations extracted from English and Ukrainian texts, the work introduces a structural validation and visualization layer based on Association Rule Mining and analytics based on graphs. The objective is to expose statistically meaningful relationships between thematic clusters and to visualize these dependencies in a transparent, interpretable manner.*

***Keywords*** *– multilingual knowledge graphs, association rule mining, thematic validation, graph visualization, cybersecurity, cross-lingual analysis.*

**Introduction**. In multilingual text mining, particularly in cybersecurity intelligence, themes often emerge through unsupervised clustering of entity-relation graphs [1, 2]. While such clustering reveals broad thematic fields, it rarely explains how these themes interact or reinforce one another across languages. Understanding these interdependencies is essential for interpretability, validation, and analytical transparency [3]. This research addresses that need by transforming co-occurrence patterns among clusters into statistically validated association rules [4] and mapping them as a Thematic Relationship Graph (TRG)**.**

**Methods and Materials**. The framework structure is presented in Fig. 1. Multilingual cybersecurity corpora were processed using large language models (LLMs) to extract named entities and semantic relations. The resulting English and Ukrainian graphs were independently clustered using community-detection algorithms to uncover thematic groupings such as phishing, advanced persistent threat (APT) opera-



* науковий керівник

tions, or critical-infrastructure defense. The bilingual graph was constructed based on results of cross-alignment between English and Ukrainian graphs while ensuring the semantic similarity between the cluster names [5]. To evaluate structural dependencies among clusters, Association Rule Mining was applied to the bilingual dataset. The co-occurrence of entities across clusters was analyzed using Support, Confidence, and Lift to identify statistically significant associations.

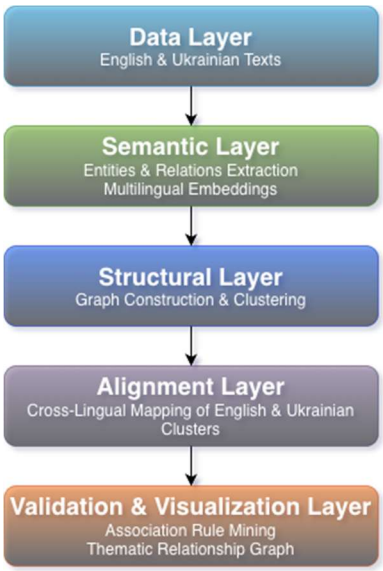


Figure 1. The Framework Architecture.

These rules will be visualized in the TRG, where:

- Nodes will represent thematic clusters or semantic fields.
- Edges will denote validated associations weighted by Lift, indicating the strength and relevance of thematic proximity.

This representation highlights both intra- and cross-lingual linkages, revealing how topics propagate across the English and Ukrainian information spaces.

**Results**. The developed visualization is expected to reveal complex structural relationships among multilingual cybersecurity themes. The TRG should highlight:

- Core thematic hubs (clusters that co-occur frequently).
- Cross-lingual dependencies where semantically related clusters align between English and Ukrainian corpora.
- Emergent thematic intersections, identifying new or hybrid areas, e.g. the overlap between technical vulnerabilities and geopolitical disinformation.



* науковий керівник

The resulting graphs are anticipated to show a layered and weighted structure where the strength of thematic connections reflects underlying data-driven relationships rather than subjective assumptions.

**Conclusions**. With the outcomes of this research, we expect to gain the ability to evaluate and visualize thematic coherence across languages. The TRG provides an intuitive yet rigorous mechanism for exploring how multilingual topics interact, evolve, or diverge. This framework will make it easier to:

- Validate discovered clusters using measurable association metrics;
- Support adaptive monitoring of multilingual threat intelligence or disinformation trends;
- Facilitate integration with further analytic tools such as trend detection, anomaly monitoring, or interactive dashboards.

Overall, this work moves toward a hybrid, interpretable, and scalable methodology for multilingual knowledge-graph analysis - combining neural discovery, symbolic reasoning, and visual exploration to enhance understanding of complex information environments.

**References**

[1] R. K. Mishra, H. Raj, S. Urolagin, J. A. A. Jothi, and N. Nawaz (2022). Cluster-Based Knowledge Graph and Entity-Relation Representation on Tourism Economical Sentiments. *Applied Sciences*, 12(16), 8105. DOI: 10.3390/app12168105

[2] L. Ding, S. Zhou, J. Xiao, and J. Han (2024). Automated Construction of Theme-Specific Knowledge Graphs. arXiv preprint, arXiv:2404.19146. DOI: 10.48550/arXiv.2404.19146

[3] M. Naeem, W. Ozuem, K. Howell, and S. Ranfagni (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International* Journal of Qualitative Methods, 22, 16094069231205789. DOI: 10.1177/16094069231205789

[4] Y. Li, M. Thomas, and K.-M. Osei-Bryson (2014). Using Association Rules Mining to Facilitate Qualitative Data Analysis in Theory Building. In K.-M. Osei-



* науковий керівник

Bryson and O. Ngwenyama (Eds.), Advances in Research Methods for Information Systems Research: Data Mining, Data Envelopment Analysis, Value Focused Thinking (pp. 79–91). Boston, MA: Springer US. DOI: 10.1007/978-1-4614-9463-8_7

[5] B. Pavlyshenko and M. Stasiuk (2025). Semantic Similarity Analysis Using Transformer-Based Sentence Embeddings. *Electronics and Information Technologies / Електроніка та інформаційні технології*, (30), 43–58. DOI: 10.30970/eli.30.4

## ***Візуалізація та валідація багатомовних тематичних структур у графах знань***

*Микола Стасюк, Богдан Павлишенко*

*У дослідженні представлено методологію для дослідження та валідації тематичних структур у багатомовних текстах у кібербезпеці. Робота ґрунтується на двомовних графах знань, що інтегрують сутності та відносини між ними, вилучені з англомовних і україномовних текстів. Запропоновано шар структурної валідації та візуалізації, який базується на методі пошуку асоціативних правил і графовій аналітиці. Метою є виявлення статистично значущих зв'язків між тематичними кластерами та візуалізація цих залежностей.*



* науковий керівник