Using LLMs to Automate Schema Mappings for RDF Knowledge Graphs Construction

Yannis Marketakis
marketak@ics.forth.gr
Institute of Computer Science, FORTH-ICS
Computer Science Department, University of Crete
Heraklion, Crete, Greece

Milio Lintanff–Castel
milio.lintanff--castel@etudiant.univ-rennes.fr
Lannion Institute of Technology, University of Rennes
Rennes, Upper Brittany, France

Yannis Tzitzikas
tzitzik@ics.forth.gr
Institute of Computer Science, FORTH-ICS
Computer Science Department, University of Crete
Heraklion, Crete, Greece

Abstract

Knowledge Graphs (KGs) are often produced through the semantic integration of heterogeneous data sources using a unified ontology. Schema mapping languages support this process by specifying the correspondences between source schemata and target ontologies; among them, the X3ML mapping definition language is widely used in domains, such as cultural heritage and biodiversity. However, creating schema mappings remains a manual and expertise-intensive task that limits large-scale integration. This paper investigates how Large Language Models (LLMs) can assist in automating the construction of schema mappings using X3ML. We present a benchmark built from real-world X3ML mappings and evaluate multiple prompting strategies across several LLMs. The results offer insights into the feasibility of using LLMs for schema mappings generation.

ACM Reference Format:

Yannis Marketakis, Milio Lintanff–Castel, and Yannis Tzitzikas. 2026. Using LLMs to Automate Schema Mappings for RDF Knowledge Graphs Construction . In *The 41st ACM/SIGAPP Symposium on Applied Computing (SAC '26), March 23–27, 2026, Thessaloniki, Greece.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3748522.3779763

1 Introduction and Background

Semantic data integration aims to harmonize heterogeneous data sources by mapping them to a unified ontology. A key step in this process is the definition of schema mappings, that specify how source schema elements correspond to classes and properties of a target ontology. X3ML framework [7] provides a widely adopted, human-readable mapping language that supports collaborative authoring and separates mapping logic from URI generation, improving transparency and reuse.

Despite these advantages, defining schema mappings remains a labor-intensive, expertise-dependent task requiring iterative collaboration among technical, domain, and ontology specialists. As data volume and heterogeneity grow, this manual process becomes increasingly difficult to scale. Recent advances in LLMs offer a promising opportunity to streamline mapping definition: their natural language understanding and contextual reasoning capabilities can help identify semantic correspondences, propose candidate alignments, and generate schema mappings, potentially accelerating the overall integration workflow.

In this work, we investigate whether LLMs can support the automatic construction of X3ML schema mappings, targeting the ISO 21127:2023 CIDOC CRM ontology [2], a widely used standard in cultural heritage. Automating the creation of such mappings is particularly valuable because they serve as an intermediate layer between source data and RDF KGs, since they can be verified, debugged, and reused more easily compared to direct LLM-based data transformations. Using real-world mappings from existing projects, we design and evaluate several prompting strategies for generating X3ML mappings.

Research on transforming heterogeneous data into RDF has produced a variety of approaches and systems, as surveyed in [11, 10]. Extensive work also exists on schema and ontology matching. More recently, LLMs have been explored for related tasks such as entity resolution[6], entity alignment[12], and direct data-to-RDF transformation[8, 1, 3, 4, 9]. The work most closely related to ours is [5], which evaluates LLMs for generating RML mappings over a small synthetic ontology. Our work differs by targeting the widely adopted CIDOC CRM and X3ML language by performing a systematic evaluation over a benchmark of real-world mappings.

Our contributions are: (a) LLM-driven approaches for generating X3ML schema mappings, (b) a benchmark dataset of real X3ML mappings targeting CIDOC-CRM, (c) a systematic evaluation of these methods, and (d) a comparative analysis of multiple LLMs and prompting strategies.

2 Benchmark

We constructed our benchmark from schema mapping projects created in the 3M Editor deployments maintained by FORTH. 3M Editor is an online tool that supports the collaborative construction and exchange of X3ML mappings. These projects were developed and validated by domain and ontology experts, and span several real-world cultural heritage datasets. For consistency, we focus on mappings targeting the ISO 21127:2023 CIDOC CRM ontology. This curated subset serves as a gold standard for evaluating automatically generated mappings. Our final dataset includes 13 mapping projects, comprising 48 mapping domains and 301 mapping links, covering diverse entity types such as persons, locations, organizations, and events.

Formally, we shall use S to denote a *schema*, where a schema consists of a set of *nodes*, denoted by $n(S)$, and a set of *relations*

SAC '26, March 23–27, 2026, Thessaloniki, Greece

ACM ISBN 979-8-4007-2294-3/2026/03
https://doi.org/10.1145/3748522.3779763

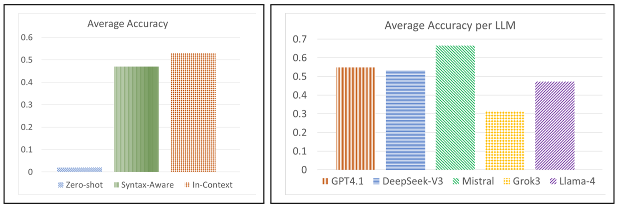

Figure 1: Average accuracy score using different prompting methods (left) and evaluated LLMs (right)

among the nodes, denoted by $r(S)$. In the context of a mapping project, let S be the source schema, and T the target schema. A mapping, from S to T, is actually a pair $M_{S,T} = (m_d, m_l)$ where m_d maps nodes of S to nodes of T, and m_l maps (relation, node)-pairs of S to (relation, node)-pairs of T. In particular, m_d is a partial function $m_d : n(S) \rightarrow n(T)$, while m_l is a partial function $m_l : r(S)\times n(S) \rightarrow r(T) \times n(T)$. Let $cont(M_{S,T})$ be the union of all the functions, i.e. $cont(M_{S,T}) = m_d \cup m_l$. We can now define the following metrics:

• **ExactMatch**. This metric checks if two mapping projects are exactly equal, thus it is a binary metric defined as: $EM(M_{AI}, M_{GS}) = (cont(M_{AI}) == cont(M_{GS}))$.

• **Accuracy.** It measures the proportion of correct mappings among all generated mappings, and it is defined as:
$A(M_{AI}, M_{GS}) = \frac{|cont(M_{AI}) \cap cont(M_{GS})|}{|cont(M_{AI})|}$

ExactMatch captures complete correctness, while Accuracy provides a more tolerant, partial-overlap assessment.

We also implemented a Python-based evaluation framework that loads gold standard mappings, generates new ones using different LLMs, and computes the metrics. It also outputs summary reports and visual results. All resources are available on GitHub[1].

3 Evaluation

We assessed several state-of-the-are LLMs for generating X3ML schema mappings: GPT-4.1, DeepSeek-V3, Mistral, Grok-3, Llamma-4. This diversity selection includes both proprietary and open-weight models, allowing us to examine how different architectures and training paradigms affect mappings quality. All models were evaluated using the same input data and the same prompts to ensure a controlled comparison.

We tested three prompting strategies that vary in the degree of syntactic and semantic guidance: (a) **zero-shot**, only the XML source data and a minimal instruction to generate X3ML mappings. (b) **Syntax-aware**, source data plus a small, domain-irrelevant but valid X3ML snippet to guide the syntax. (c) **In-Context**, source data plus a complete, domain-relevant X3ML mapping project. All prompts target the CIDOC CRM ontology and instruct the model to output valid X3ML.

We computed the accuracy metric for each model, dataset, and prompting strategy. Figure 1 (left) shows the average accuracy per prompting method; as expected, Zero-Shot achieves near-zero accuracy, while the other two are more accurate. Figure 1 (right) summarizes the average accuracy per LLM. Mistral-Medium consistently performs best, followed by GPT-4.1 and DeepSeek-V3. Grok-3 obtained the lowest accuracy. In several cases, models returned invalid X3ML files, resulting in accuracy q; these were treated as outright failures since they cannot be executed by X3ML Engine - the data transformation engine that is responsible for constructing RDF KGs, using source data and X3ML schema mappings. More detailed results can be found on GitHub.

[1]https://github.com/ymark/x3ml_comparator

4 Conclusions

In this paper, we investigated different methods that utilize LLMs for supporting the construction of schema mappings for transforming structured data. We described the construction of a benchmark that comprises schema mappings from real projects, and conducted experiments with various LLMs and prompt strategies, yielding an average accuracy of 53%. The results demonstrate that the laborious and manual process of schema definition can be substantially accelerated or partially automated. Future work will explore improved prompting strategies, including pre-submission steps designed to enhance the quality of the generated schema mappings.

Acknowledgments

This project has received funding from SemantyFish project (Horizon Europe OSCARS 1st Call for Open Science Projects and Services).

References

[1] Salvatore Carta, Alessandro Giuliani, Leonardo Piano, Alessandro Sebastian Podda, Livio Pompianu, and Sandro Gabriele Tiddia. Iterative zero-shot llm prompting for knowledge graph construction. *arXiv preprint arXiv:2307.01128*, 2023.
[2] Martin Doerr. The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine*, 24(3):75–75, 2003.
[3] Johannes Frey, Lars-Peter Meyer, Natanael Arndt, Felix Brei, and Kirill Bulert. Benchmarking the abilities of large language models for rdf knowledge graph creation and comprehension: How well do llms speak turtle? *arXiv preprint arXiv:2309.17122*, 2023.
[4] Skander Ghazzai, Daniela Grigori, Boualem Benatallah, and Raja Rebai. Harnessing gpt for data transformation tasks. In *2024 IEEE International Conference on Web Services (ICWS)*, pages 1329–1334. IEEE, 2024.
[5] Marvin Hofer, Johannes Frey, and Erhard Rahm. Towards self-configuring knowledge graph construction pipelines using llms-a case study with rml. In *Fifth International Workshop on Knowledge Graph Construction@ ESWC2024*, volume 3718, 2024.
[6] Huahang Li, Longyu Feng, Shuangyin Li, Fei Hao, Chen Jason Zhang, and Yuanfeng Song. On leveraging large language models for enhancing entity resolution: a cost-efficient approach. *arXiv preprint arXiv:2401.03426*, 2024.
[7] Yannis Marketakis, Nikos Minadakis, Haridimos Kondylakis, Konstantina Konsolaki, Georgios Samaritakis, Maria Theodoridou, Giorgos Flouris, and Martin Doerr. X3ml mapping framework for information integration in cultural heritage and beyond. *International Journal on Digital Libraries*, 18:301–319, 2017.
[8] Anmol Nayak and Hari Prasad Timmapathini. Llm2kb: Constructing knowledge bases using instruction tuned context aware large language models. *arXiv preprint arXiv:2308.13207*, 2023.
[9] Sanaz Saki Norouzi, Adrita Barua, Antrea Christou, Nikita Gautam, Andrew Eells, Pascal Hitzler, and Cogan Shimizu. Ontology population using llms. *arXiv preprint arXiv:2411.01612*, 2024.
[10] Vetle Ryen, Ahmet Soylu, and Dumitru Roman. Building semantic knowledge graphs from (semi-) structured data: a review. *Future Internet*, 14(5):129, 2022.
[11] Dylan Van Assche, Thomas Delva, Gerald Haesendonck, Pieter Heyvaert, Ben De Meester, and Anastasia Dimou. Declarative rdf graph generation from heterogeneous (semi-) structured data: A systematic literature review. *Journal of Web Semantics*, 75:100753, 2023.
[12] Runhao Zhao, Weixin Zeng, Jiuyang Tang, Yawen Li, Guanhua Ye, Junping Du, and Xiang Zhao. Towards unsupervised entity alignment for highly heterogeneous knowledge graphs. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pages 3792–3806. IEEE Computer Society, 2025.