

# Enriching Taxonomies using Large Language Models

Zeinab Ghamlouch<sup>\*</sup> and Mehwish Alam<sup>\*\*</sup>

<sup>a</sup>Télécom Paris, Institut Polytechnique de Paris, France  
ORCID (Mehwish Alam): <https://orcid.org/0000-0002-7867-6612>

**Abstract.** Taxonomies play a vital role in structuring and categorizing information across domains. However, many existing taxonomies suffer from limited coverage and outdated or ambiguous nodes, reducing their effectiveness in knowledge retrieval. To address this, we present Taxoria, a novel taxonomy enrichment pipeline that leverages Large Language Models (LLMs) to enhance a given taxonomy. Unlike approaches that extract internal LLM taxonomies, Taxoria uses an existing taxonomy as a seed and prompts an LLM to propose candidate nodes for enrichment. These candidates are then validated to mitigate hallucinations and ensure semantic relevance before integration. The final output includes an enriched taxonomy with provenance tracking and visualization of the final merged taxonomy for analysis.

## 1 Introduction

Taxonomies play a crucial role in structuring and hierarchically categorizing information for better organization and clarity. They are widely used across various domains including web data, e-commerce, digital libraries, general purpose knowledge graphs (such as YAGO [6], DBpedia, Wikidata), and biomedical domain (such as MeSH, SNOMED-CT, etc.) to support efficient information retrieval, semantic reasoning, and knowledge management. However, many existing taxonomies, such as Schema.org, exhibit limitations in their coverage and granularity. These include missing, ambiguous or outdated nodes. Such gaps hinder the effectiveness of taxonomy-driven systems and reduce their ability to reflect the evolving nature of knowledge. Enriching a taxonomy entails enhancing its structural depth and contextual relevance by introducing new, semantically meaningful categories without disrupting its original hierarchy.

Many existing taxonomies are either derived from textual data or manually built in response to user needs. Large Language Models (LLMs) on the other hand, are trained on large amount of data on the web, storing plethora of knowledge. GPTKB [3] is one of the knowledge graphs which given a seed entity extracts the knowledge of an LLM about that entity in the form of triples. It also extracts the taxonomy of the LLMs which the authors claim to be close to the taxonomy of Schema.org. In contrast, while GPTKB explicitly extracts the internal taxonomy of an LLM, our approach uses a seed taxonomy to guide the LLM’s output.

In this work, we propose a novel taxonomy enrichment pipeline, Taxoria, which enables live enrichment of a given taxonomy using the inherent knowledge of an LLM. It leverages the existing taxonomy as a seed for generating candidate nodes for enrichment using

zero-shot prompting. These suggestions are post-processed and validated using various techniques to mitigate hallucinations and over-generations. After validation, the node is merged into the original taxonomy with the provenance information. Additionally, it also provides visualization of the merged taxonomy allowing deeper analysis. The live demo is available online<sup>1</sup>, the source code is available through Github<sup>2</sup>, along with the video demonstration<sup>3</sup>.

This paper is structured as follows: Section 2 places the tool in reference to existing studies. Section 3 gives the overall architecture of Taxoria while Section 4 discusses the use-case of real-world taxonomies. Finally, Section 5 highlights the limitations and discusses the future directions.

## 2 Related Work

Traditional approaches to taxonomy enrichment, such as manual curation, rule-based systems, and statistical models face critical challenges. YAGO4.5 [6] performs manual taxonomy refinement with expert intervention using Schema.org as the base taxonomy, however, manual enrichment is resource-intensive. Rule-based methods lack adaptability and generalizations [1], and statistical techniques often fail to capture deeper semantic relationships. On the other hand, LLMs capture large body of knowledge including taxonomic knowledge. Various studies have utilized LLMs for taxonomy expansion [4] and taxonomy refinement [5]. In [5], the authors use a combination of graph mining techniques and LLMs, where LLM is used for verifying the links between the nodes. Taxoria is close to the task of taxonomy expansion, however, it uses the internal knowledge of an LLM for the expansion.

## 3 Overall Architecture

Figure 1, shows the overall architecture of Taxoria. This automated pipeline focuses on enriching the taxonomy based on the internal taxonomy of an LLM. The enrichment process involves the following key steps: starting with a seed taxonomy, the system prompts the LLM for extracting the internal taxonomy of different open-source LLMs retaining only relevant information.

**Taxonomy Traversal Strategy.** We employ two primary taxonomy traversal strategies:

- **Breadth-First Search (BFS).** All nodes at a given level are expanded before proceeding to the next level.

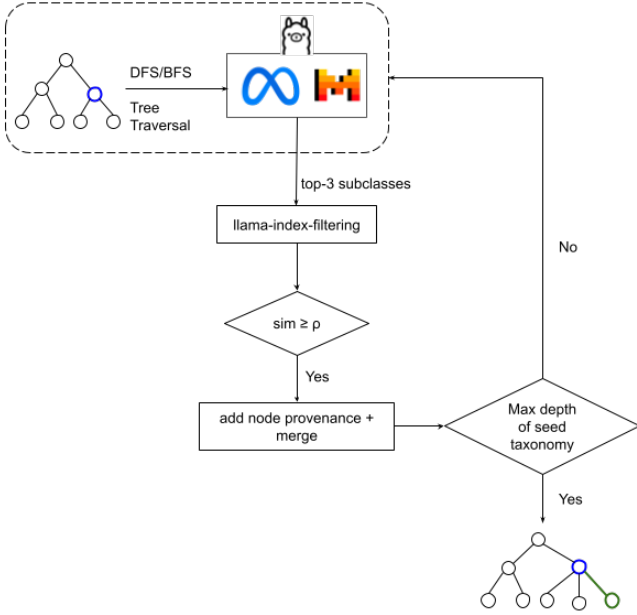
<sup>\*</sup> Corresponding Author. Email: [zeinab.ghamlouch@ip-paris.fr](mailto:zeinab.ghamlouch@ip-paris.fr).

<sup>\*\*</sup> Email: [mehwish.alam@telecom-paris.fr](mailto:mehwish.alam@telecom-paris.fr)

<sup>1</sup> <https://taxoria.r2.enst.fr/>

<sup>2</sup> <https://github.com/zeinabGhamlouch/Taxoria>

<sup>3</sup> <https://www.youtube.com/watch?v=7-9Hrg-awGo>



**Figure 1:** Overall architecture of Taxoria. The blue node represents the current node in the seed taxonomy for which the direct children are generated and the green node represents the newly added node after the merging process.

- **Depth-First Search (DFS).** A complete branch is expanded before moving to adjacent branches.

**Prompting LLM.** Zero-shot prompting is used for generating the classes from an LLM using seed node from a taxonomy traversed using one of the previously defined strategies. For a given node only the top-3 direct children were generated which use the internal scoring of an LLM with a constraint to generate an output in a JSON format.

**Mitigating Over-generations LLM.** Despite their capabilities, LLMs may produce suggestions that are irrelevant, redundant, or inconsistent due to their tendencies to hallucinate. To address this, we apply the following strategies:

- **Defining Stopping Criteria.** To prevent over-generation and avoid deeper taxonomy, we define stopping criteria based on the maximum depth of the original taxonomy. Newly generated nodes are restricted to extend the hierarchy by at most one additional level beyond the existing maximum depth.
- **Filtering based on Semantic Similarity.** We compute the semantic similarity between a generated node and its parent using Word2Vec embeddings. A node is merged into the taxonomy if its cosine similarity to the parent node is greater than a certain threshold  $\rho$  (here  $\rho \geq 0.9$  within the range  $[0,1]$ ). For computing the semantic similarity Word2Vec and LlamaIndex were used for ensuring n-gram level as well as contextual level similarity respectively.

### 3.1 Merging process

Figure 2 shows the overall merging process. It integrates new nodes generated by the LLM into the existing taxonomy while avoiding redundancy and preserving the hierarchical structure. The process consists of the following steps:

**Node Generation.** New candidate nodes are generated by the LLM based on the current state of the taxonomy. These nodes are proposed as direct children of existing nodes or as deeper subtrees, depending on the tree traversal strategy used.

**Validation of Node Relevance.** Each generated node is evaluated using the filtering mechanisms described previously to ensure conceptual relevance, avoid redundancy, and maintain structural coherence.

**Merging Nodes.** The system handles node integration based on whether the generated node already exists in the taxonomy:

- If the node already exists with the same name, the system recursively compares and merges their children. The original children are preserved, and any new children are appended, avoiding duplication.
- If no matching node is found, the new node is directly added as a child to the appropriate parent node in the taxonomy.

**Adding Class Provenance.** To distinguish between original and newly generated nodes, we introduce a **source key** attribute for tracking the provenance of each node:

- **Original Nodes.** All nodes that were part of the initial taxonomy retain a source key with the value “original-taxonomy”, indicating that they originate from the original taxonomy.
- **Newly Merged Nodes.** All nodes generated and added through the enrichment process are assigned a source key with the value “llm-generated”, signifying that they were derived from the LLM-based pipeline.

This explicit tracking mechanism ensures transparency and traceability of enriched content.

### 3.2 Implementation Details

Taxoria implements open source LLMs, i.e., LLAMA 3, LLAMA 3.2, and Mistral. Each of these LLMs are chosen based on their characteristics, i.e.,

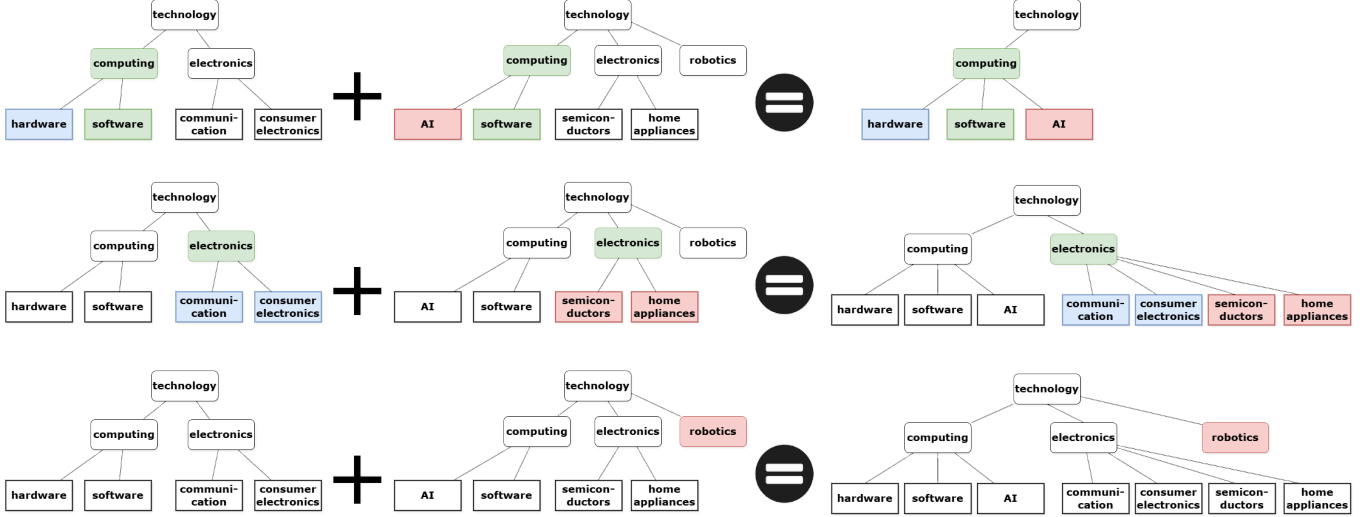
- **LLaMA 3 & 3.2** are chosen for their strong language generation performance and low-latency while generating responses.
- **Mistral** is selected for its compact model size and efficient inference speed.

Ollama<sup>4</sup> is chosen to run the LLMs on the backend. This choice is motivated by the fact that, the smaller LLMs can be run locally offering control over the environment and ensuring data privacy which is one of the most important concerns in various industry/domain oriented applications. Additionally, it also provides access to a number of LLMs and not only the ones chosen in this work allowing flexibility while building applications. Taxoria uses NVIDIA GTX 1080 GPU, LLMs with higher number of parameters can be used upon the availability of higher computational resources. The use of Ollama also allows to run Taxoria on local machine.

## 4 Use-case and Impact

This section discusses the application of Taxoria on real-taxonomies leading to a discussion on the broader impact of the tool.

<sup>4</sup> <https://ollama.com/>



**Figure 2:** This figure illustrates an example of merging two hierarchical taxonomies in the technology domain. Nodes common to both input taxonomies are shown in **green**, while nodes unique to each taxonomy are marked in **blue** and **red**, respectively.

**Real-World Taxonomy Enrichment.** We tested Taxoria in the real setting with real world taxonomies. These taxonomies span diverse domains, including general knowledge (Schema.org<sup>5</sup>), e-commerce (eBay), and geographical information (GeoNames) [7]. Schema.org is tailored for structured metadata on the web, eBay for hierarchical product categorization, and GeoNames for geographic entity classification. Table 1 shows the statistics of the taxonomies and the number of newly LLM-generated classes.

**Analysis.** While enriching the previously described taxonomies, we encountered predictions which were irrelevant or were entities and not the classes. However, these strategies are not directly integrated into the demo in order to ensure it functions within limited computational resources.

- **LLM-Based Relevance Filtering.** In this scenario, we use LLM-as-a-judge for evaluating the relevance of a generated node with respect to its parent. The node is retained only if its relevance score meets or exceeds the average similarity observed in the original taxonomy.
- **Validation against an External Knowledge Graph.** In some cases, LLMs generate instances (entities) of a class, e.g., *Paris* which is an instance of a class *City*. During this validation process, we use external knowledge graph such as Wikidata [8] to verify if the generated node is an instance or a class. For doing so, Wikidata is queried to check the existence of “instance of” (P31) property. If the property is not associated to generated node, the node is retained otherwise it is discarded.

Taxonomy	# Classes	Max Depth	# New Classes	New Max Depth
Schema.org	1,143	6	4216	7
eBay	596	4	1106	5
GeoNames	690	3	1138	4

**Table 1:** Statistics of the Taxonomy. The # new classes are reported only for Llama 3.2.

**Impact.** Taxonomies defined in general purpose knowledge graphs such as DBpedia, Wikidata, and YAGO have various limitations. The

taxonomy of DBpedia is semi-automatically generated which is incomplete. Wikidata’s taxonomy is often disorganized as a result of its crowd-sourced contributions. However, the taxonomy of YAGO has evolved over time. It started with the use of WordNet hierarchy, which then evolved to using Wikidata taxonomy. The taxonomy of the current version of, YAGO 4.5, is manually curated based on Schema.org. This tool can further be used for creating richer taxonomies for these knowledge graphs. The live enrichment of this taxonomy can also serve as an agent for these knowledge graphs, enabling dynamic integration of new information.

Taxonomies related to a certain domain such as biomedicine are vast and complex, often maintained separately. Merging tools can help integrate taxonomies across sources, while LLMs can be leveraged to enrich them with up-to-date biomedical knowledge, especially where scientific field is evolving fast and it is not possible to manually update the taxonomy.

## 5 Discussion and Future Directions

While applying Taxoria on the real-world taxonomies, we observed that at a certain level of the taxonomy instead of generating classes as the subclasses of the node from the seed taxonomy, the LLM starts generating instances of those classes. In such a scenario we decided to use external knowledge graphs such as Wikidata for grounding the LLM generation by using the “instance of” relation using SPARQL queries. However, this led to various issues, i.e., given that Wikidata is a semi-automatically generated and is a crowd sourced effort, it contains many inconsistencies and erroneous type information. As a future direction, we would extend the tool to simultaneously perform entity type prediction [2] for a certain domain related to seed taxonomy.

The current visualization support of Taxoria does not supporty edits. In future, this could be extended to allow human evaluation within the tool by allowing live user edits. Further extensions would support tasks such as taxonomy induction, expansion, and refinement. Additionally, robust validation strategies will be explored to ensure the accuracy and reliability of the merged taxonomies, including both automated evaluation metrics and human-in-the-loop approaches.

<sup>5</sup> <https://schema.org/>

## References

- [1] H. Angermann. Taxomulti: Rule-based expert system to customize product taxonomies for multi-channel e-commerce. *SN Comput. Sci.*, 3(2): 177, 2022. doi: 10.1007/S42979-022-01070-8. URL <https://doi.org/10.1007/s42979-022-01070-8>.
- [2] R. Biswas, J. Portisch, H. Paulheim, H. Sack, and M. Alam. Entity type prediction leveraging graph walks and entity descriptions. In *The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings*, volume 13489 of *Lecture Notes in Computer Science*, pages 392–410. Springer, 2022.
- [3] Y. Hu, S. Ghosh, T. Nguyen, and S. Razniewski. GPTKB: building very large knowledge bases from language models. *CoRR*, abs/2411.04920, 2024. doi: 10.48550/ARXIV.2411.04920. URL <https://doi.org/10.48550/arXiv.2411.04920>.
- [4] S. Mishra, U. Sudev, and T. Chakraborty. Flame: Self-supervised low-resource taxonomy expansion using large language models. *ACM Trans. Intell. Syst. Technol.*, 2024. ISSN 2157-6904. doi: 10.1145/3709007. URL <https://doi.org/10.1145/3709007>.
- [5] Y. Peng, T. Bonald, and M. Alam. Refining wikidata taxonomy using large language models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM*, 2024. doi: 10.1145/3627673.3679156. URL <https://doi.org/10.1145/3627673.3679156>.
- [6] F. M. Suchanek, M. Alam, T. Bonald, L. Chen, P. Paris, and J. Soria. YAGO 4.5: A large and clean knowledge base with a rich taxonomy. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR*, 2024.
- [7] Y. Sun, X. Hao, K. Sun, Y. Xu, X. Yang, X. L. Dong, N. Tang, and L. Chen. Are large language models a good replacement of taxonomies? *Proc. VLDB Endow.*, 17(11):2919–2932, 2024.
- [8] D. Vrandečić and M. Krötzsch. Wikidata: a free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85, 2014. doi: 10.1145/2629489. URL <https://doi.org/10.1145/2629489>.