

Refining Wikidata Taxonomy using Large Language Models

Yiwen Peng

yiwen.peng@telecom-paris.fr

Télécom Paris

Institut Polytechnique de Paris

Palaiseau, France

Thomas Bonald

thomas.bonald@telecom-paris.fr

Télécom Paris

Institut Polytechnique de Paris

Palaiseau, France

Mehwish Alam

mehwish.alam@telecom-paris.fr

Télécom Paris

Institut Polytechnique de Paris

Palaiseau, France

Abstract

Due to its collaborative nature, Wikidata is known to have a complex taxonomy, with recurrent issues like the ambiguity between instances and classes, the inaccuracy of some taxonomic paths, the presence of cycles, and the high level of redundancy across classes. Manual efforts to clean up this taxonomy are time-consuming and prone to errors or subjective decisions. We present WiKC, a new version of Wikidata taxonomy cleaned automatically using a combination of Large Language Models (LLMs) and graph mining techniques. Operations on the taxonomy, such as cutting links or merging classes, are performed with the help of zero-shot prompting on an open-source LLM. The quality of the refined taxonomy is evaluated from both intrinsic and extrinsic perspectives, on a task of entity typing for the latter, showing the practical interest of WiKC.

CCS Concepts

• Information systems → Data structures.

Keywords

Knowledge Graphs, Large Language Model, Graph Mining

ACM Reference Format:

Yiwen Peng, Thomas Bonald, and Mehwish Alam. 2024. Refining Wikidata Taxonomy using Large Language Models. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3627673.3679156>

1 Introduction

Wikidata is a general-purpose Knowledge Base (KB) maintained by a large community of contributors. As a collaborative project, Wikidata faces several challenges, including the ambiguity, inconsistency, redundancy, and complexity of its taxonomy. Ambiguity arises from the confusion between instances and classes. For example, *scientist* (Q901) is both an instance of *profession* (Q28640) and a subclass of *person* (Q215627). Inconsistency here refers to the inaccuracy of some taxonomic paths. For instance, *city* (Q515) is inaccurately classified as a subclass of *mathematical object* (Q246672) through the following taxonomic path: *city* (Q515) → *spatial entity* (Q58416391) → *geometric object* (Q123410745) → *mathematical object* (Q246672). Redundancy is also prevalent with classes like

human (Q5) and *person* (Q215627) coexisting, where one would suffice. The complexity of the taxonomy is another major issue. The taxonomy of Wikidata has a depth of 20, contains many cycles, like *axiom* (Q17736) → *first principle* (Q536351) → *principle* (Q211364) → *axiom* (Q17736), and transitive links, like *airport* (Q1248784) → *aerodrome* (Q62447) → *station* (Q12819564) and *airport* (Q1248784) → *station* (Q12819564). Additionally, only 4% of the 4 million classes are instantiated, with many lacking labels and descriptions.

In this paper, we propose an approach for refining Wikidata taxonomy and thus addressing some of the above issues. Unlike YAGO 4.5 [19], where the upper taxonomy of Wikidata is manually mapped to Schema.org¹ while the lower taxonomy remains unchanged, we adopt an automated approach for refining the whole taxonomy, based on a combination of Large Language Models (LLMs) and graph mining techniques. Specifically, we use a zero-shot prompt on each link of the taxonomy, asking the LLM to predict one of the following relations: *subclassOf*, *superclassOf*, *equivalent*, *irrelevant*, or *none*. Given different predictions, we decide whether to cut the link, merge the classes, or keep the classes and their link unchanged. This yields WiKC, a cleaned version of Wikidata taxonomy, which we make publicly available². The quality of WiKC is then evaluated from both intrinsic and extrinsic perspectives. For extrinsic evaluation, we design a judge LLM on the task of entity typing to compare WiKC with the original taxonomy of Wikidata. For the sake of reproducibility, both steps (refinement and evaluation) are based on an open-source LLM.

The paper is structured as follows. Section 2 covers the related work, Section 3 describes our approach for taxonomy refinement, Section 4 presents the evaluation and Section 5 concludes the paper.

2 Related Work

General-purpose knowledge bases. Wikidata [21] is the largest open general-purpose KB, maintained by a large community of contributors. Due to its collaborative nature, Wikidata is known to have a complex taxonomy, including errors, redundancies and inconsistencies [6, 18]. Cleaning this taxonomy is the main objective of our work. Other general-purpose KBs include DBpedia [3] and YAGO 4.5 [19]. DBpedia is a multilingual KB automatically extracted from Wikipedia and more recently, Wikidata. Its ontology covers a wide range of concepts but suffers from inconsistencies [1, 17] due to its reliance on Wikipedia and the prioritization of coverage over precision. YAGO 4.5 is based on Wikidata, with a manual mapping of the upper taxonomy to Schema.org, providing a clean upper-level taxonomy designed by human experts. In this paper, we propose an automatic approach for refining the Wikidata taxonomy, without requiring any human expertise or subjective decision.



This work is licensed under a Creative Commons Attribution International 4.0 License.

CIKM '24, October 21–25, 2024, Boise, ID, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0436-9/24/10

<https://doi.org/10.1145/3627673.3679156>

¹<https://schema.org>

²<https://github.com/peng-yiwen/WiKC>

Taxonomy refinement, taxonomy induction. Taxonomy refinement is the task of updating an existing taxonomy while maintaining its structure. Previous methods are either domain-specific [15] or depend on lexical structures of existing hierarchies [16]. Recently, more advanced approaches have incorporated word embeddings into taxonomy refinement. For instance, hyperbolic embeddings are used in [2] to detect outliers in a domain-specific taxonomy. In [14], a hierarchical semantic similarity metric is used to select better embeddings and then refine a taxonomy.

Few studies have explored the use of LLMs in taxonomy refinement. Instead, LLMs have been applied to the closely related task of taxonomy induction [4, 8, 11, 23], which derives a taxonomy from scratch given entities extracted from text. For example, an approach called Chain-of-Layer is used in [23] to select relevant entities. A zero-shot knowledge-agnostic strategy is used in [7] for constructing the upper levels of a taxonomy. In [11], a concept hierarchy is generated for a given domain starting from a seed concept. The reliability of LLMs in hierarchical structure discovery is demonstrated in [20] for common knowledge graphs, including Schema.org. While these studies show the potential of LLMs in taxonomy induction, the automatic refinement of a large common taxonomy like that of Wikidata remains an open challenge.

3 Approach

In our work, we use the *truthy* version of Wikidata³, which contains the best non-deprecated rank for each property.

3.1 Taxonomy Extraction

In principle, the taxonomy of Wikidata is defined by the *subclassOf* (P279) property. In practice, this property is often confused with the *instanceOf* (P31) property by contributors to Wikidata, requiring some work to extract the actual taxonomy.

Instance or class? We extract instances and classes using the *instanceOf* and *subclassOf* properties, respectively, giving priority to the *instanceOf* property if both appear. For example, the entity *hydrogen* (Q556), which is both an instance of *chemical element* (Q11344) and a subclass of *energetic material* (Q5376832), is considered as an instance, not a class. However, some exceptions must be taken into account. For example, the entity *company* (Q783794), which is an instance of *type of organisation* (Q17197366) and a subclass of *organization* (Q43229), should be considered as a class. The difference with the previous example is that the entity *type of organisation* is, in fact, a *metaclass* (Q19478619), i.e., a class which has instances that are all themselves classes. Given that, the entity *company* should indeed be considered as a class, not an instance.

In our case, we consider that an entity is a meta-class if it is an instance of either *metaclass* (Q19478619) or *second-order class* (Q24017414) that meets the following criteria: (1) Its label contains a keyword like *type*, *class*, *style*, *genre*, *form*, *occupation*, *profession*, *category*, *classification*, (2) Its label does not contain a preposition, which corresponds to very specific classes, nor the keyword *property*, which refers to classes of properties.

We also exclude *BFO class* (Q124711104) from the meta-classes to avoid external ontologies. We finally obtain 434 meta-classes and approximately 1.7M classes (either an instance of a meta-class or

an entity that has the *subclassOf* property and not the *instanceOf* property; meta-classes are excluded). The class *product* (Q2424752) was manually added because it is an important class, e.g., one of the 11 top-level classes of Schema.org, filtered out by the previous process as an instance of *economic concept* (Q29028649).

Graph construction. Among all entities declared as classes by the previous process, we only keep those with a label. We build a directed graph between these 1.6M classes using the *subclassOf* property. We then explore the taxonomy from the root class *entity* (Q35120) by depth-first-search (DFS) and break any link that would create a cycle. During traversal, we also bypass any class without a description, as most of these classes are either redundant or overly specific. For example, the class *award for best screenplay* (Q96474700) is a subclass of *award for best book* (Q105810971), which is itself a subclass of *literary award* (Q378427); here we bypass *award for best book* due to the absence of description in Wikidata.

Filtering. We exclude the class *scholarly article* (Q13442814) and all its successors in the graph, as the addition of scholarly articles to Wikidata has been controversial⁴ and would contribute more than 40M entities. We also remove classes that do not have instances (there are 252k of them). In addition, we eliminate top-level classes (subclasses of root class *entity*) that do not have subclasses themselves, such as *unidentified entity* (Q120725535) and *named entity* (Q25047676). These classes are either derived from external ontologies or are too general to be useful in our taxonomy.

After completing the aforementioned steps, we obtain an acyclic taxonomy with approximately 40k classes and 53k links, where each class has a label and a description.

3.2 Taxonomy Refinement

We refine the taxonomy to address issues like redundancy and inconsistency. For instance, the class *city or town* (Q27676416) is a subclass of *city or town* (Q7930989) (redundancy) and a transitive subclass of *mathematical object* (Q246672) (inconsistency). For this, we prompt an LLM (see details in §3.3) to analyze each link of the graph and predict the correct semantic relation from the following ones: *subclassOf*, *superclassOf*, *equivalent*, *irrelevant*, and *none*. The *superclassOf* prediction is used to potentially reverse the link direction. Given these results, obtained for each link of the graph, we apply the following steps sequentially: (1) **Cut** irrelevant links; (2) **Resolve** reversed links; (3) **Reduce** transitive links; (4) **Merge** equivalent classes; (5) **Rewire** links upon confirmation by LLM and (6) **Filter** out non-informative and rare classes. The evolution of the subgraph corresponding to the paths from *city or town* (Q7930989) to *entity* (Q35120) is shown in Figure 1, after each refining step. The details for each step are described below.

Cut. Any link that is predicted as *irrelevant* or *none* is cut if the corresponding classes remain connected to the root class *entity* after the cut, or if the disconnected subgraph after the cut has at most 3 nodes (in which case the corresponding classes are also removed from the taxonomy). Links are considered in order of distance from the root class *entity*. In Figure 1b for instance, *city or town* is no longer a transitive subclass of *mathematical object*, because the link from *spatial entity* to *mathematical object* is cut.

³Data dump dated March 22, 2024, and 949GB after unpacking.

⁴<https://www.mail-archive.com/wikidata@lists.wikimedia.org/msg06716.html>

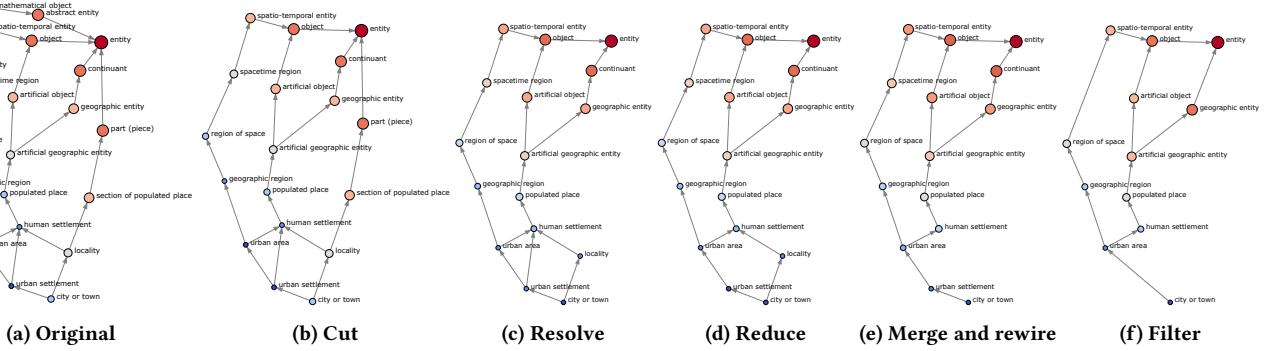


Figure 1: Taxonomy from *city or town* (Q7930989) to *entity* (Q35120) after each step of the refinement.

Resolve. When the link prediction is reversed, i.e., class A is predicted as a *superclassOf* class B instead of a *subclassOf* class B , we merge these classes or cut their link depending on the connectivity of class A in the graph: if class A is a subclass of other classes than B , the link between classes A and B is cut; otherwise, classes A and B are merged. Merging is reasonable in this resolving step as reversed links are usually caused by similar classes, making it hard to decide which one is a subclass of the other. In Figure 1c, the link from *locality* to *section of populated place* is cut as *locality* has another superclass, *human settlement*.

Merge. We merge classes that have exactly the same label or that are predicted as equivalent by the LLM. Specifically, if class A is predicted as *equivalent* to class B instead of a *subclassOf* class B , class A is merged into class B . The subclasses of class A are then relinked to class B as its subclasses. A transitive reduction is performed after each merging operation to prevent the introduction of transitive links and cycles in the taxonomy. In Figure 1e for instance, the class *locality*, defined as a *place of human settlement*, is merged into class *human settlement* as these two classes are predicted as equivalent.

Rewire. After the previous merge step, we inspect the potential subclass links between class B (the former superclass of class A), and the other superclasses of class A , if any. Here we use the same LLM and prompt to check these links and only accept those correctly predicted as *subclassOf*. Our experiments show that, out of such 280 potential subclass links, 118 are correctly predicted by the LLM. For instance, after the merge of *tweezers* (Q192504) into *forceps* (Q1378235), *forceps* becomes a subclass of *hand tool* (Q2578402), another superclass of *tweezers*, after confirmation of the LLM.

Filter. In this final step, to further overcome redundancy issues, we remove recursively *non-informative* classes (classes with only one superclass, one subclass, and without direct instances), *rare* classes (classes with at most one instance including both direct and transitive instances, or without a Wikipedia page⁵), and *specific* top-level classes (top-level classes whose subclasses are all linked to other deeper level classes)⁶. The remaining classes are reconnected

⁵For classes without any Wikipedia page, we consider only classes with a depth higher than 3, as the upper-level classes might be too abstract to be described in Wikipedia (e.g., *artificial object* (Q16686448)).

⁶For instance, *testbed* (Q1318674) is a *specific* top-level class. The subclass of *testbed*, *elevator test tower* (Q1689156), is also a subclass of a deeper level class *tower* (Q12518), and *testbed* is too specific to be directly linked to the root class *entity* (Q35120).

respecting the previous taxonomy structure if no transitive links are created. In Figure 1f for instance, the non-informative class *spacetime region* is removed and the class *region of space* becomes a subclass of *spatio-temporal entity*; the rare class *urban settlement* is also removed (due to the absence of a Wikipedia page) and the class *city or town* becomes a subclass of *urban area*.

After refining the taxonomy of Wikidata, we obtain a new acyclic taxonomy called WiKC, with approximately 17k classes and 20k links, without any cycles and transitive links.

3.3 Large Language Models

For the sake of reproducibility, we use the open-source LLM *Mixtral-8x7B-Instruct-v0.1*⁷, with a temperature set to zero to get deterministic results. We formulate a prompt to enable the LLM to generate answers from the context, including class labels and descriptions. Inspired by the chain-of-thought [22], which bridges the reasoning gap between input and answer, we add an explanation part before the answer to ensure careful analysis before predicting the semantic relation. The corresponding prompt is shown below.

```
%Instructions:  
You are a linguistic expert who understands the semantic relationships between concepts. Your task is to determine the most appropriate semantic relation between two provided concepts based on the available labels and descriptions. The potential relations are: "subclass of", "superclass of", "equivalent to", "irrelevant to", or "None" if none applies. You should select exclusively from these relation options and not introduce other relationships.  
Please structure your response as follows:  
Response:::  
Explanation: (your analysis of the semantic relation between two concepts).  
Answer: (state the relation explicitly, e.g. "ConceptA is [relation] ConceptB")  
  
%User Message:  
Examine the relation between the following two concepts, each described below:  
* ConceptA: labeled as "{child_label}", described as "{child_description}."  
* ConceptB: labeled as "{parent_label}", described as "{parent_description}."  
Response:::
```

4 Evaluation

We assess the quality of WiKC from intrinsic and extrinsic perspectives. Data, code, prompts, and resources are all available online⁸.

Intrinsic evaluation. We verified the inclusion of the 40 upper-level classes of YAGO 4.5⁹ in WiKC. There are only two exceptions: *yago:Gender*, which can be represented by the property *sex* or *gender*

⁷<https://huggingface.co/mistralai/Mixtral-8x7B-Instruct-v0.1>

⁸<https://github.com/peng-yiwen/WiKC>

⁹<https://yago-knowledge.org/data/yago4.5/design-document.pdf>

(P21), and *schema:Taxon*, which is a specific class of the biological domain. Following YAGO 4.5 [19], we evaluate WiKC in terms of three key criteria: complexity, conciseness, and understandability (fraction of classes having labels and descriptions). The statistics are given in Table 1. As expected, WiKC is much simpler and much more concise than Wikidata taxonomy. Compared to WiKC, Wikidata taxonomy has a factor higher than 200 in the number of classes, and a factor higher than 10 in the average number of paths from an instance to the root class *entity* (Q35120).

Table 1: Quality Measures

Criterion	Metric	Wikidata	WiKC
Complexity	Classes	4.1M	17k
	Top-level classes	38	16
	Links	4.8M	20k
	Depth	20	13
Conciseness	Average paths to root	37	2.9
	Cycles	35	0
	Redundant links	500k	0
Understandability	Classes without instances	3.9M	0
	Labels and descriptions	78%	100%

Extrinsic evaluation. We further evaluate WiKC on a task of entity typing, i.e., the prediction of the classes of an entity. This is a crucial task for various downstream tasks, like entity alignment [13] or entity linking [12]. Our evaluation includes the direct classes of each instance as well as their ancestors in the taxonomy, in order to assess the inconsistency of some taxonomic paths.

We collect instances from the Wikidata dump based on the *instanceOf* (P31) or *occupation* (P106) relations, excluding scholarly articles (described in Section §3.1), ensuring each instance has a label, a description, and an English Wikipedia page (resulting in 7M instances). We retype these instances using WiKC by assigning instances to their nearest classes in the taxonomy. To avoid class distribution imbalance (e.g., the class *person* can have 2.6M cumulative instances), we limit each class to 1000 instances and randomly sample 100k samples overall, resulting in nearly 1M type statements per taxonomy. We design a judge LLM¹⁰ to verify the accuracy of type statements based on the context provided by an instance. For example, given the context: **Paris* is described as the capital of France*, the LLM judges if the statement **Paris* is a [city or town], which means 'large human settlement'* is True or False. In this case, the class within brackets can be any ancestor of *city* (Q515), the direct class of *Paris* (Q90).

Table 2 demonstrates the accuracy of entity typing across different depths of the taxonomy on Wikidata and WiKC, where depth refers to the shortest distance from the root class *entity*. The results show that WiKC consistently outperforms Wikidata across all depth ranges. WiKC shows significant accuracy gains at deeper levels (depth 10 or more), suggesting that WiKC has resolved many inconsistency issues in the lower levels of the Wikidata taxonomy. The fact that accuracy is higher at a deeper level (depth 5 or more) compared to a shallow level on WiKC can be explained by the fact that more specific types are easier for LLMs to judge. For example,

¹⁰We use the same model *Mixtral-8x7B-Instruct-v0.1*.

it is easier to classify *Motokazu Mori* (Q75688679) as a *poet* (Q49757) (depth 9) than as a *corporate body* (Q106668099) (depth 2).

Table 2: Accuracy of entity typing on Wikidata and WiKC.

Depth	[0, 5)	[5, 10)	[10, ∞)	Macro
Wikidata	41%	47%	37%	43%
WiKC	67%	76%	75%	70%

Discussion. We here discuss some limits of our work.

(1) Problems with the LLM. The LLM might hallucinate by producing responses in conflict with the input prompt. For example, when checking the link *coke plant* (Q905318) \rightarrow *coke* (Q192795), the LLM generates a response where *coke* (Q192795) is a subclass of *coke plant's product*, creating a new class instead of respecting the input. It should generate *none* if no appropriate relation is found between the two classes, rather than hallucinating a non-existent class in Wikidata. Additionally, the LLM can exhibit inconsistency between the explanation part and the answer part, although this happens rarely. For instance, in the explanation part, *input device* (Q864114) is analyzed as one type of *physical interface* (Q64830866) that *specifically provides data and signals to an information processing system*. But the answer incorrectly orders the classes, stating that *physical interface* is a subclass of *input device*.

(2) Applications to downstream tasks. Even though we perform both intrinsic and extrinsic evaluations of WiKC, we cut away large parts of the Wikidata taxonomy (from about 4M to 17k classes). Therefore, it is also important to assess the knowledge coverage of WiKC and its usefulness for downstream tasks.

5 Conclusion

In this paper, we propose WiKC, a cleaned version of Wikidata taxonomy, generated by an automated process combining zero-shot prompting on an open-source LLM and graph mining approaches. The objective is to address several known limits of Wikidata taxonomy, such as inaccurate taxonomic paths, redundancy across classes, complexity, and ambiguity between instances and classes. Our approach consists of cutting irrelevant links, resolving reversed links, reducing transitive links, merging equivalent classes, rewiring links upon reconfirmation of LLM, and filtering out non-informative or rare classes. The experimental results show the improved accuracy and conciseness of WiKC compared to the original taxonomy of Wikidata. In addition, we provide a mapping file from WiKC to Wikidata, encouraging the reuse of WiKC in various downstream tasks, such as entity recognition [10], entity linking [5] and entity summarization [9], to further validate its reliability and its coverage of general knowledge.

For future work, we consider directions for exploring other open-source LLMs to clean and evaluate taxonomies based on our proposed pipeline, and investigating the trustworthiness of these LLMs in the taxonomy refinement task. It is also valuable to share this approach with the Wikidata community to further check its feasibility and help alleviate the burden of manual taxonomy cleaning¹¹.

¹¹https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology/Cleaning_Task_Force

References

- [1] David Abián, Francesco Guerra, J Martínez-Romanos, and Raquel Trillo-Lado. 2018. Wikidata and DBpedia: a comparative study. In *Semantic Keyword-Based Search on Structured Data Sources: Third International KEYSTONE Conference, IKC 2017, Gdańsk, Poland, September 11–12, 2017, Revised Selected Papers and COST Action IC1302 Reports 3*. Springer, 142–154.
- [2] Rami Aly, Shantanu Acharya, Alexander Ossa, Arne Köhn, Chris Biemann, and Alexander Panchenko. 2019. Every child should have parents: a taxonomy refinement algorithm based on hyperbolic term embeddings. *arXiv preprint arXiv:1906.02002* (2019).
- [3] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A nucleus for a web of open data. In *international semantic web conference*. Springer, 722–735.
- [4] Hamed Babaei Giglou, Jennifer D’Souza, and Sören Auer. 2023. LLMs4OL: Large language models for ontology learning. In *International Semantic Web Conference*. Springer, 408–427.
- [5] Edoardo Barba, Luigi Procopio, and Roberto Navigli. 2022. ExtEnD: Extractive entity disambiguation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2478–2488.
- [6] Freddy Brasileiro, João Paulo A Almeida, Vitorio A Carvalho, and Giancarlo Guizzardi. 2016. Applying a multi-level modeling theory to assess taxonomic hierarchies in Wikidata. In *Proceedings of the 25th international conference companion on World Wide Web*. 975–980.
- [7] Salvatore Carta, Alessandro Giuliani, Leonardo Piano, Alessandro Sebastian Poddà, Livio Pompianu, and Sandro Gabriele Tiddia. 2023. Iterative zero-shot llm prompting for knowledge graph construction. *arXiv preprint arXiv:2307.01128* (2023).
- [8] Boqi Chen, Fandi Yi, and Dániel Varró. 2023. Prompting or Fine-tuning? A Comparative Study of Large Language Models for Taxonomy Construction. In *2023 ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. IEEE, 588–596.
- [9] Andrew Chisholm, Will Radford, and Ben Hachey. 2017. Learning to generate one-sentence biographies from Wikidata. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 633–642.
- [10] Besnik Fetahu, Zhiyu Chen, Sudipta Kar, Oleg Rokhlenko, and Shervin Malmasi. 2023. MultiCoNER V2: a Large Multilingual dataset for Fine-grained and Noisy Named Entity Recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 2027–2051.
- [11] Maurice Funk, Simon Hosemann, Jean Christoph Jung, and Carsten Lutz. 2023. Towards Ontology Construction with Language Models. *arXiv preprint arXiv:2309.09898* (2023).
- [12] Nitish Gupta, Sameer Singh, and Dan Roth. 2017. Entity Linking via Joint Encoding of Types, Descriptions, and Context. In *Conference on Empirical Methods in Natural Language Processing*. <https://api.semanticscholar.org/CorpusID:28784495>
- [13] Hongren Huang, Chen Li, Xutan Peng, Lifang He, Shu Guo, Hao Peng, Lihong Wang, and Jianxin Li. 2021. Cross-knowledge-graph entity alignment via relation prediction. *Knowl. Based Syst.* 240 (2021), 107813. <https://api.semanticscholar.org/CorpusID:245304392>
- [14] Lorenzo Malandri, Fabio Mercurio, Mario Mezzanica, and Navid Nobani. 2021. Taxoref: Embeddings evaluation for ai-driven taxonomy refinement. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 612–627.
- [15] Timothy D O’Hara, Andrew F Hugall, Ben Thuy, Sabine Stöhr, and Alexander V Martynov. 2017. Restructuring higher taxonomy using broad-scale phylogenomics: the living Ophiuroidea. *Molecular phylogenetics and evolution* 107 (2017), 415–430.
- [16] Simone Paolo Ponzetto, Roberto Navigli, et al. 2009. Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia.. In *IJCAI*, Vol. 9. 2083–2088.
- [17] Zhaohua Sheng, Xin Wang, Hong Shi, and Zhiyong Feng. 2012. Checking and handling inconsistency of DBpedia. In *Web Information Systems and Mining: International Conference, WISM 2012, Chengdu, China, October 26–28, 2012. Proceedings*. Springer, 480–488.
- [18] Kartik Shenoy, Filip Ilievski, Daniel Garijo, Daniel Schwabe, and Pedro Szekely. 2022. A study of the quality of Wikidata. *Journal of Web Semantics* 72 (2022), 100679.
- [19] Fabian M Suchanek, Mehwish Alam, Thomas Bonald, Lihu Chen, Pierre-Henri Paris, and Jules Soria. 2024. 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*. 131–140.
- [20] Yushi Sun, Hao Xin, Kai Sun, Yifan Ethan Xu, Xiao Yang, Xin Luna Dong, Nan Tang, and Lei Chen. 2024. Are Large Language Models a Good Replacement of Taxonomies? *arXiv preprint arXiv:2406.11131* (2024).
- [21] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledge base. *Commun. ACM* 57, 10 (2014), 78–85.
- [22] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.
- [23] Qingkai Zeng, Yuyang Bai, Zhaoxuan Tan, Shangbin Feng, Zhenwen Liang, Zhihan Zhang, and Meng Jiang. 2024. Chain-of-Layer: Iteratively Prompting Large Language Models for Taxonomy Induction from Limited Examples. *arXiv preprint arXiv:2402.07386* (2024).