Answering Complex Questions Using Text-to-SPARQL Semantic Parsing

Farhad Ghassemi
Stanford University
Computer Science Department
farhadg@stanford.edu

Melvin Orichi Socana
Stanford University
Electrical Engineering Department
msorichi@stanford.edu

Abstract

In the pursuit of enhancing the factuality of Large Language Models (LLMs), this paper presents an approach to entity linking within the context of semantic parsing for Wikidata. Recognizing the limitations of existing entity linkers in handling short, real-world questions with sparse context, we introduce a pre-processing step that utilizes the generative capabilities of GPT to provide enriched contextual descriptions for entities. This method significantly improves the performance of ReFinED Ayoola et al. (2022), a state-of-the-art (SOTA) entity linker for Wikidata, as evidenced by a marked increase in the F1 score. Leveraging this new entity linking process, we further fine-tune an LLM to generate SPARQL queries, anticipating a corresponding boost in the accuracy of the final query output. Our approach showcases the power of integrating LLMs with structured knowledge bases, paving the way for more reliable and verifiable question-answering systems. The implications of this advancement extend beyond question answering, offering potential improvements in any application requiring precise entity disambiguation.

1 Introduction

The advent of Large Language Models (LLMs) has revolutionized the field of natural language processing, offering unprecedented capabilities in generating human-like text and answering complex questions. However, despite their linguistic prowess, LLMs are prone to producing factually incorrect information—a phenomenon often attributed to their reliance on patterns in data rather than verifiable knowledge. To mitigate this issue, grounding LLMs in structured knowledge bases like Wikidata has emerged as a promising solution, ensuring that the generated content is anchored in factual reality.

The motivation behind our research stems from the observation that while grounding LLMs can enhance their factuality, the process is heavily dependent on the accuracy of entity linking. Entity linking serves as the critical bridge between unstructured text and structured data, enabling the correct mapping of entities mentioned in natural language queries to their corresponding identifiers in a knowledge base. However, existing entity linking systems struggle with queries that are short and lack context, often failing to identify relevant entities or disambiguate between similarly named ones (Ling et al., 2015). This limitation is particularly pronounced in real-world question-answering benchmarks, where users pose queries without providing full entity names or additional context.

To address this challenge, we introduce an entity linking strategy that leverages the generative capabilities of GPT before employing ReFinED (Ayoola et al., 2022), an entity-linker developed for Wikidata. By prompting GPT to generate a description and contextual information for each entity within a query, we significantly enrich the input for ReFinED, allowing for more accurate entity recognition. This pre-processing step harnesses the innate understanding of LLMs to provide a nuanced context that traditional entity linkers are not designed to infer.

Our main contribution lies in the substantial improvement of entity linking accuracy within the semantic parsing framework for Wikidata, presented in Xu et al. (2023). By integrating our enhanced entity linking method with an LLM fine-tuned to generate SPARQL queries, we not only achieve a significant increase in the F1 score for correct entity recognition but an improved overall query output accuracy. The implications of this advancement are profound, as it not only bolsters the reliability of question-answering systems but also demonstrates the potential of LLMs to act as intermediaries between natural language and structured databases.

In summary, our work presents a synergistic approach that combines the strengths of generative LLMs with the precision of structured knowledge bases. Through this integration, we pave the way for more factual and dependable applications across various domains where accurate entity linking is paramount.

2 Related Work

The field of semantic parsing over large knowledge bases has seen significant advancements with the introduction of Large Language Models (LLMs) and their application in question-answering systems. Our work builds upon and contributes to this growing body of research, particularly focusing on the intersection of LLMs, entity linking, and semantic parsing for Wikidata.

2.1 Fine-tuned LLMs Know More, Hallucinate Less

This seminal work by Xu et al. (2023) laid the foundation for our research by demonstrating the potential of fine-tuning LLMs to enhance their understanding of formal query languages and their application in semantic parsing. By leveraging a few-shot learning approach, the authors were able to fine-tune a LLaMA (Touvron et al., 2023) with a training set, thereby improving the learnability of SPARQL queries. Their methodology replaced the IDs of properties and domains with their unique names and tolerated errors in entity linking by accepting mentions in the queries as entities. This approach established a strong baseline for answer accuracy on their newly contributed WikiWebQuestions benchmark and surpassed the state-of-the-art for the QALD-7 Wikidata set in F1 score and exact matches (EM) (Xu et al., 2023).

2.2 ReFinED and Entity Linking

Entity linking is the task of assigning a unique identity to entities mentioned in text and further disambiguating similarly named entities. RefinED, as presented by Ayoola et al. (2022), represents a significant leap forward in entity linking, providing a robust system capable of linking entities within user queries to their unique identifiers in knowledge bases like Wikidata. However, as identified in our research, ReFinED faces challenges when dealing with short, context-poor queries typical of real-world question-answering scenarios. Our work addresses these challenges by introducing a pre-processing step that utilizes GPT to generate enriched contextual descriptions for entities before they are fed into ReFinED. This innovation significantly improves the precision of entity recognition

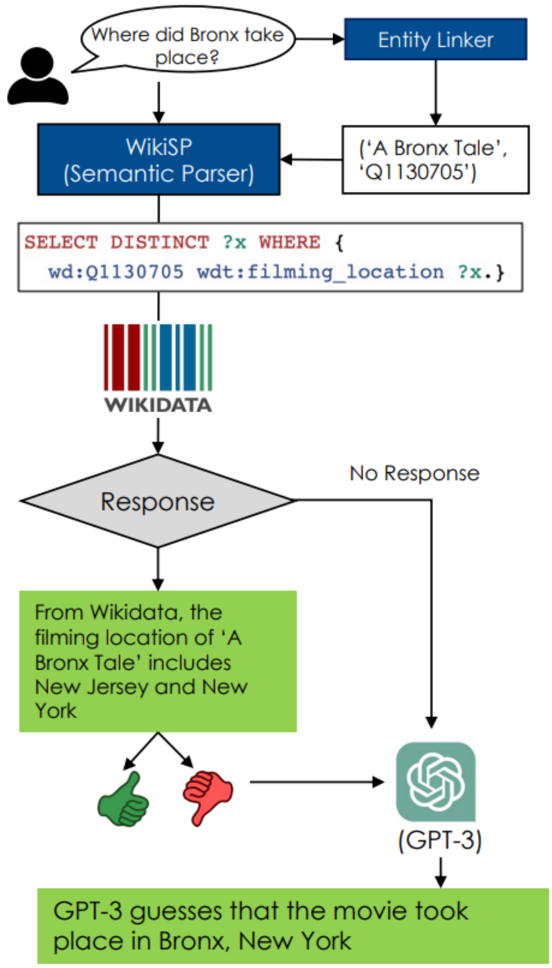

Figure 1: A summary of WikiSP: WikiSP utilizes an entity linker to connect user query entities to their specific IDs in Wikidata. For instance, "A Bronx Tale" might be linked to the entity ID "Q1130705". The results from the query and entity linker are then used as inputs for the WikiSP semantic parser. Retrieved directly from Xu et al. (2023)

and, by extension, the accuracy of the semantic parser output.

Our research works off of these pivotal studies, extending their methodologies and addressing some of their limitations. By integrating an enhanced entity linking process with a fine-tuned LLM for generating SPARQL queries, we not only improve upon the existing benchmarks but also open new avenues for future research in creating more reliable and verifiable question-answering systems grounded in structured knowledge bases.

3 Method

In our work to refine the accuracy of semantic parsing for Wikidata, our research introduces a

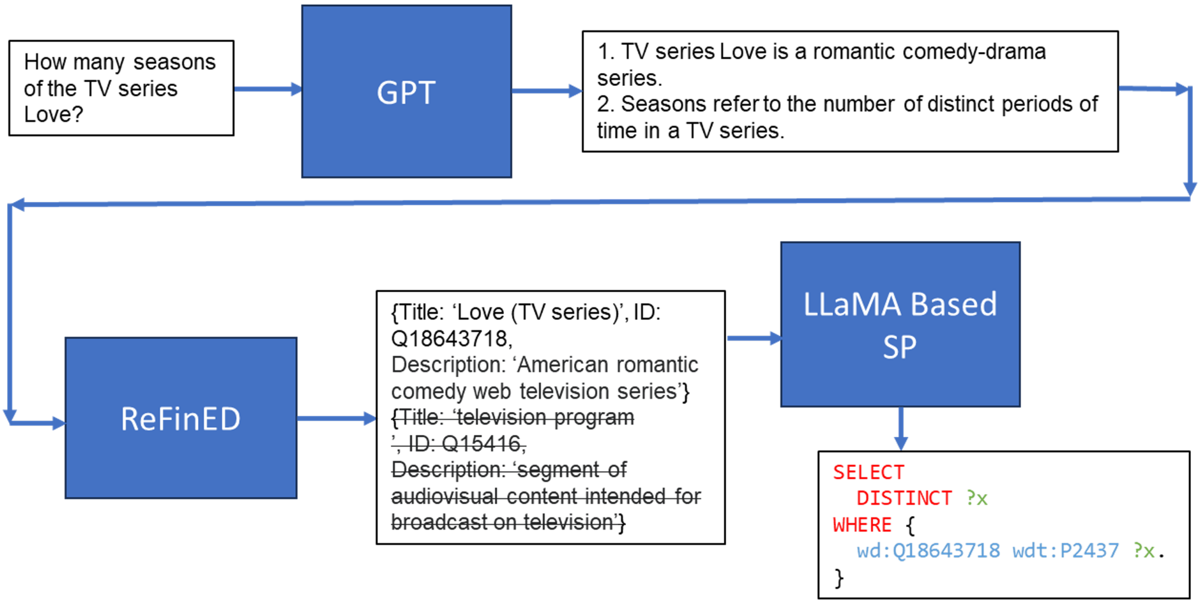

Figure 2: An illustration of our proposed entity-linking pipeline. Before being passed to ReFinED, an LLM is tasked with entity-detection and disambiguation and generates relevant descriptions. The original input sentence in addition to the new descriptions are then input to ReFinED to perform scoring. Finally, the question and entities are inputted into the semantic parser to obtain a SPARQL query to retrieve relevant results. Note that in this example, without the additional step of generating entity descriptions, ReFinED by itself fails to find the relevant entities.

two-tiered approach that significantly enhances the entity linking process. This methodology is predicated on the insight that the generative capabilities and powerful knowledge of LLMs can be harnessed to provide a richer context for entities mentioned in user queries, thereby improving the precision of subsequent entity disambiguation by ReFinED.

3.1 Contextual Enrichment with Generative LLMs

The first core idea involves leveraging a generative LLM—specifically, GPT—to preemptively interpret and expand upon the entities present within a query. By prompting GPT to produce a description and additional context for each entity, we effectively create a more detailed backdrop against which ReFinED can operate. We show our prompts in the appendices. This step is crucial in addressing the inherent limitations of short, context-poor queries that often lead to ambiguous or incorrect entity linking.

3.2 Enhanced Entity Linking with ReFinED

Upon receiving the enriched context from GPT, ReFinED then performs its entity linking task with this new context included. The additional information provided by GPT allows ReFinED to better distinguish between entities with similar names and to identify relevant entities that might otherwise be overlooked due to the brevity of the query.

3.3 Integration and Fine-tuning of LLMs for SPARQL Generation

We also fine-tune an LLM for the generation of SPARQL queries. This fine-tuning process is designed to capitalize on the improved entity linking, hypothesizing that a more precise identification of entities will naturally lead to an increase in the accuracy of the generated SPARQL queries.

Our methodology is simple yet powerful, addressing a critical bottleneck in the question-answering pipeline and setting a new standard for integrating structured knowledge bases with LLMs. The elegance of our approach lies in its ability to significantly improve upon existing benchmarks with minimal complexity, demonstrating that strategic enhancements to key components can yield substantial advancements in overall system performance.

4 Experimental Results

We show the experimental results we obtained in this section. First, we show results relevant to the entity linking; including the F1 score and the numbers of predicted entities between just using ReFinED for entity linking against our pipeline in

Entity Linker	Dataset	#Exact Matches	#Gold Entities	#Predicted Entities
ReFinED	Compmix	1383	1877	1714
ReFinED+GPT-3.5	Compmix	1516	1877	1866
ReFinED	WikiWebQuestions	400	563	466
ReFinED+GPT-3.5	WikiWebQuestions	426	563	480

Table 1: The table delineates the entity linking performance of the ReFinED pipeline against the enhanced ReFinED+GPT-3.5 pipeline, showcasing the latter's increased number of exact matches and predicted entities across the Compmix dev dataset (1680 questions) and WikiWebQuestions dev dataset (454 questions). We surmise that these improvements likely indicate higher precision and recall, demonstrating the efficacy of incorporating GPT-3.5 into the entity linking process.

Entity Linker	Dataset	F1 Score
ReFinED	Compmix	0.63
ReFinED+GPT-3.5	Compmix	**0.80**
ReFinED	WikiWebQuestions	0.73
ReFinED+GPT-3.5	WikiWebQuestions	**0.83**

Table 2: Comparison of average F1 scores for entity linking using ReFinED and the enhanced ReFinED+GPT-3.5 pipeline across two datasets, Compmix dev set (1680 Questions) and WikiWebQuestions dev set (454 Questions). The results highlight the significant performance gains achieved by integrating GPT-3.5 into the ReFinED framework, with bold figures indicating the improved scores.

Table 1 and Table 2. We show results on WikiWebQuestions (Xu et al., 2023) and Compmix (Christmann et al., 2023).

Lastly, after training a LLaMA to perform semantic parsing and output SPARQL queries, we show our improved results against the results reported in the original WikiSP paper (Xu et al., 2023). We evaluate our model with two different answer accuracy metrics, which are the same metrics reported in the WikiSP paper: (1) exact match (EM): the percentage of examples where the answers of the predicted SPARQL exactly match the gold answers, and (2) Macro F1 score (F1): the average F1 score for answers of each example. These evaluation results are shown in Table 3.

	Exact Match	F1 Score
WikiSP Baseline	0.756	0.769
Our Method	**0.786**	**0.802**

Table 3: We compare the performance of our semantic parsing method, which utilizes a trained LLaMA model for generating SPARQL queries, against the reported results of the WikiSP baseline.

Our method demonstrates a slight improvement in both exact match and macro F1 score metrics, with an exact match score of **0.786** and an F1 score of **0.802**, surpassing the baseline scores of 0.756 and 0.769 respectively. These results underscore the effectiveness of our approach in accurately interpreting and answering complex queries.

5 Learnings & Future Work

Throughout the course of this research, we have gleaned several key insights that have both immediate and long-term implications for the field of semantic parsing and entity linking. Our work has underscored the importance of context in entity recognition, demonstrating that even small enhancements in the pre-processing phase can lead to significant improvements in overall system performance.

Learnings:

1. A nuanced approach to entity linking that incorporates additional context can dramatically reduce the rate of disambiguation errors, particularly in queries that are inherently ambiguous or sparse in detail.

2. The generative power of LLMs like GPT can be effectively utilized beyond direct question-answering tasks, serving as a valuable tool for context generation and enrichment.

3. Fine-tuning LLMs with a focus on downstream tasks such as SPARQL query generation can yield better results when preceded by a robust entity linking process.

Future Work: Building on these learnings, our future work will aim to explore several avenues.

1. Expansion of our methodology to other structured knowledge bases beyond Wikidata, examining its applicability and effectiveness across diverse datasets.

2. Development of a more granular error analysis framework to pinpoint specific areas where our enhanced entity linking process could be further refined.

3. Applying our pipeline on other datasets and benchmarks to observe the results. These could include question answering benchmarks as shown in (Wang, 2022) or datasets commonly used to test and benchmark the performance of entity linking system such as the AIDA CoNLL-YAGO dataset presented in Hoffart et al. (2011).

6 Conclusion

In conclusion, our research has made a significant contribution to improving the enity-linking functionality of WikiSP and possibly other question-answering systems grounded in large knowledge bases like Wikidata. By innovatively leveraging the generative capabilities of LLMs to enrich the context for entity linking, we have addressed a critical challenge that has long hindered the accuracy of semantic parsers. Our methodology demonstrates that strategic enhancements to the entity linking process can lead to substantial improvements in the generation of SPARQL queries, thereby increasing the factuality and reliability of answers provided by LLMs.

The success of our approach is quantified not only by the marked increase in the F1 score for correct entity recognition but also by the anticipated improvement in the accuracy of SPARQL query outputs. This advancement represents a meaningful step forward in our quest to create more intelligent and trustworthy AI systems that can seamlessly interact with structured knowledge bases.

As we look to the future, we are encouraged by the potential applications of our work and are hopeful that it will further advance the space.

7 Acknowledgement

We would like to express our sincere appreciation to Doctor Monica Lam, Shicheng (George) Liu, and Liza Pertseva for their invaluable contributions to the creation and refinement of this paper. Their expertise, insights, and unwavering support were instrumental in the development of our research. Additionally, we extend our gratitude to the writers of the WikiSP paper, Xu et al. (2023), for their work.

References

Tom Ayoola, Shubhi Tyagi, Joseph Fisher, Christos Christodoulopoulos, and Andrea Pierleoni. 2022. ReFinED: An efficient zero-shot-capable approach to end-to-end entity linking. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Track*, pages 209–220, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Philipp Christmann, Rishiraj Saha Roy, and Gerhard Weikum. 2023. Compmix: A benchmark for heterogeneous question answering. *arXiv preprint arXiv:2306.12235*.

Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust disambiguation of named entities in text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 782–792, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Xiao Ling, Sameer Singh, and Daniel S. Weld. 2015. Design challenges for entity linking. *Transactions of the Association for Computational Linguistics*, 3:315–328.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Zhen Wang. 2022. Modern question answering datasets and benchmarks: A survey.

Silei Xu, Shicheng Liu, Theo Culhane, Elizaveta Pertseva, Meng-Hsi Wu, Sina J. Semnani, and Monica S. Lam. 2023. Fine-tuned llms know more, hallucinate less with few-shot sequence-to-sequence semantic parsing over wikidata.

A Appendix

A.1 Prompts

We show the prompt sent to GPT-3.5 to perform preliminary entity recognition and disambiguation. We start by providing examples to GPT and at the end ask the model to list the entities and their descriptions for the last question.

```
[{'role': 'system', 'content': 'You are a helpful assistant.'},

{'role': 'user', 'content': "You are a named entity recognition
   and entity disambiguation system. You are given a question and
    you need to list all entities in the question with a brief
   description for each entity. Each description should be max 10
    words. Here are some examples:

Question: what year lebron james came to the nba?
Answer:
1. LeBron James is American basketball player (born 1984)
2. National Basketball Association is North American professional
    sports league

Question: what form of government was practiced in sparta?
Answer:
1. Sparta is city-state in ancient Greece

{'role': 'user', 'content': "List the entities and their
   descriptions for this question:
Question: Where is the best place to vacation in the Dominican
   Republic?
Answer:"}]
```

Figure 3: Prompt example to instruct GPT-3.5 to provide the found entities and generate description for them for a given question. In the full prompt, 12 examples are provided to GPT.

The subsequent GPT response then provides an answer to the question: **'Where is the best place to vacation in the Dominican Republic?'**

```
GPT-Response:
1.Dominican Republic is island sovereign state in the Caribbean Sea.
2.Tourist attraction is place of interest where tourists visit.
```

This prompt is then provided to ReFinED along with the original input sentence for scoring.

```
ReFinED Input:
Where is the best place to vacation in the Dominican Republic?
1.Dominican Republic is island sovereign state in the Caribbean Sea.
2.Tourist attraction is place of interest where tourists visit.

ReFinED Response:
1.{Title: ‘Dominican Republic’, ID: Q786}
2.{Title: ‘Tourist Attraction’, ID: Q570116}
```

Lastly, the information is input into a semantic parser to output a SPARQL query. We use a finetuned LLaMA-2.

```
SELECT
    DISTINCT
?x WHERE {
    ?x wdt:P31/wdt:P279*wd:Q570116;
        wdt:P17 wd:Q786. }
```

This SPARQL query can be input into the Wikidata Query Service to observe the results; a list of locations in the Dominican Republic.