

Semantic Parsing for Conversational Question Answering over Knowledge Graphs

Laura Perez-Beltrachini

ANLP 23-11-2023

Joint work with Parag Jain¹, Emilio Monti², Mirella Lapata¹

¹ University of Edinburgh, ² Amazon Alexa

(Some slides from Mirella Lapata)



**UK Research
and Innovation**

Natural Language User Interfaces

Androutsopoulos et al. (1995)



CONVERSATION MACHINE (Green et al., 1959); BASEBALL (Green et al., 1961).

Natural Language User Interfaces

Androutsopoulos et al. (1995)

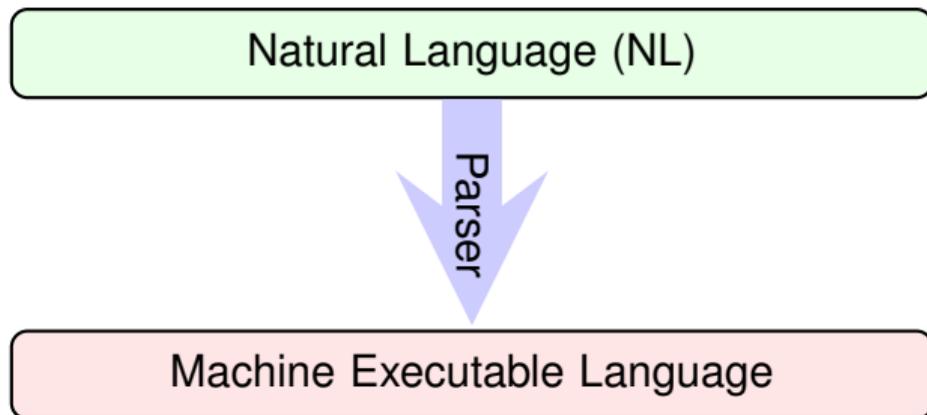
Semantic parsing: users interact with computers in human language!



CONVERSATION MACHINE (Green et al., 1959); BASEBALL (Green et al., 1961).

Natural Language User Interfaces

Semantic parsing: users interact with computers in human language!



Semantic Parsing: Answering Questions with Freebase

NL Who are the male actors in Titanic?

Parser

KB $\lambda x. \exists y. \text{gender}(\text{MALE}, x) \wedge \text{cast}(\text{TITANIC}, x, y)$



Titanic

1997 · Drama film/Romance · 3h 30m

7.7/10 · [IMDb](#)

88% · [Rotten Tomatoes](#)

James Cameron's "Titanic" is an epic, action-packed romance set against the ill-fated maiden voyage of the R.M.S. Titanic; the pride and joy of the White Star Line and, at the time, the larg... [More](#)

Initial release: November 18, 1997 ([London](#))

Director: [James Cameron](#)

Featured song: [My Heart Will Go On](#)

Cast



[Leonardo DiCaprio](#)
Jack Dawson



[Kate Winslet](#)
Rose DeWitt Bukater



[Billy Zane](#)
Caledon Heckley



[Gloria Stuart](#)
Rose DeWitt Bukater



[Kathy Bates](#)
Molly Brown

Conversational User Interfaces

- More **natural** than stand-alone questions

Who was the injured player in the sports team that was champion in the tournament where Detroit Tigers participated in?

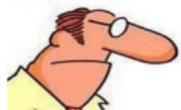
- Users carry out **exploratory searches** they learn as they search and adjust they needs as they read system answers

More complex human-machine interactions!

[Iyyer et al. (2017); Radlinski and Craswell (2017); Gao et al. (2023)]



Conversational Question Answering over Knowledge Graphs



Which tournament did Detroit Tigers participate in?



Conversational Question Answering over Knowledge Graphs



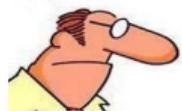
Which tournament did Detroit Tigers participate in?

1909 World Series



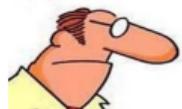
WIKIDATA

Conversational Question Answering over Knowledge Graphs



Which tournament did Detroit Tigers participate in?

1909 World Series

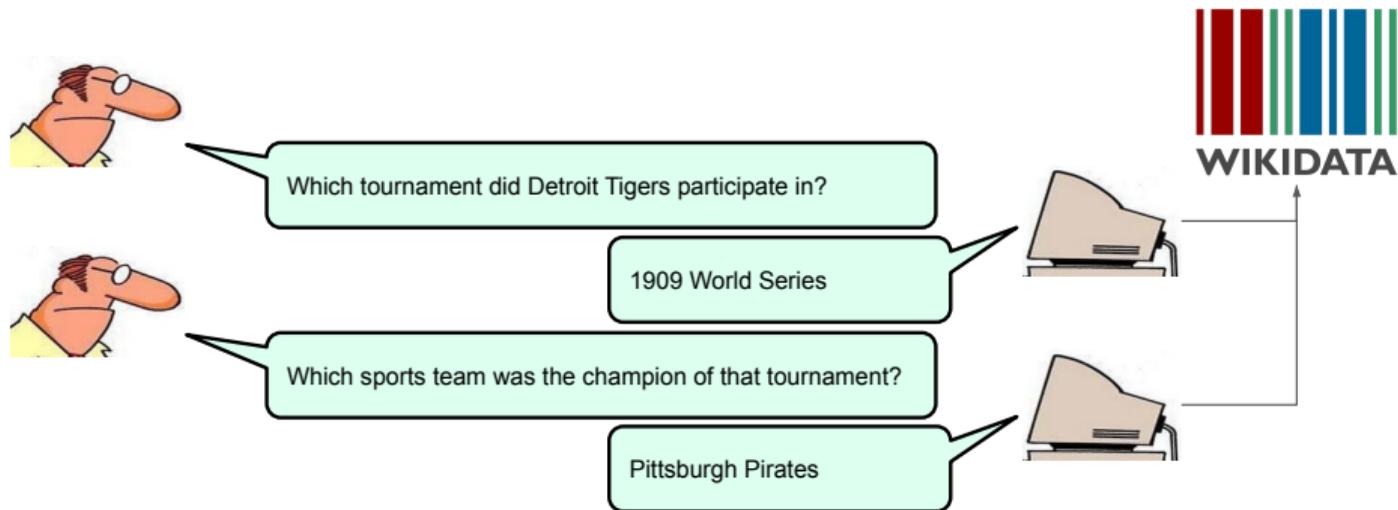


Which sports team was the champion of that tournament?

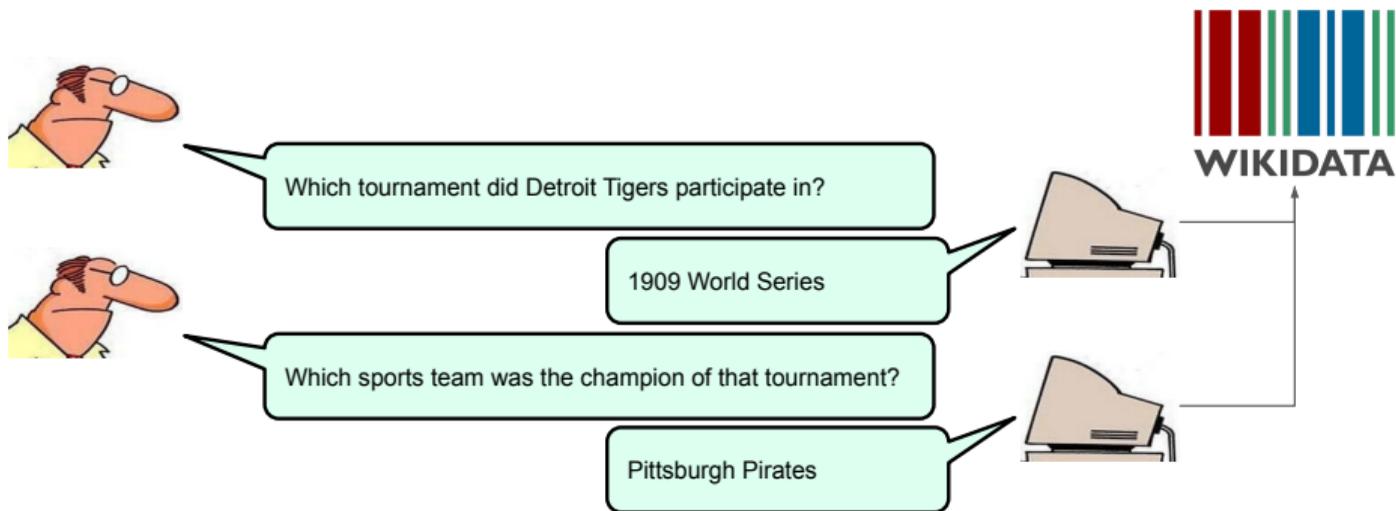


WIKIDATA

Conversational Question Answering over Knowledge Graphs

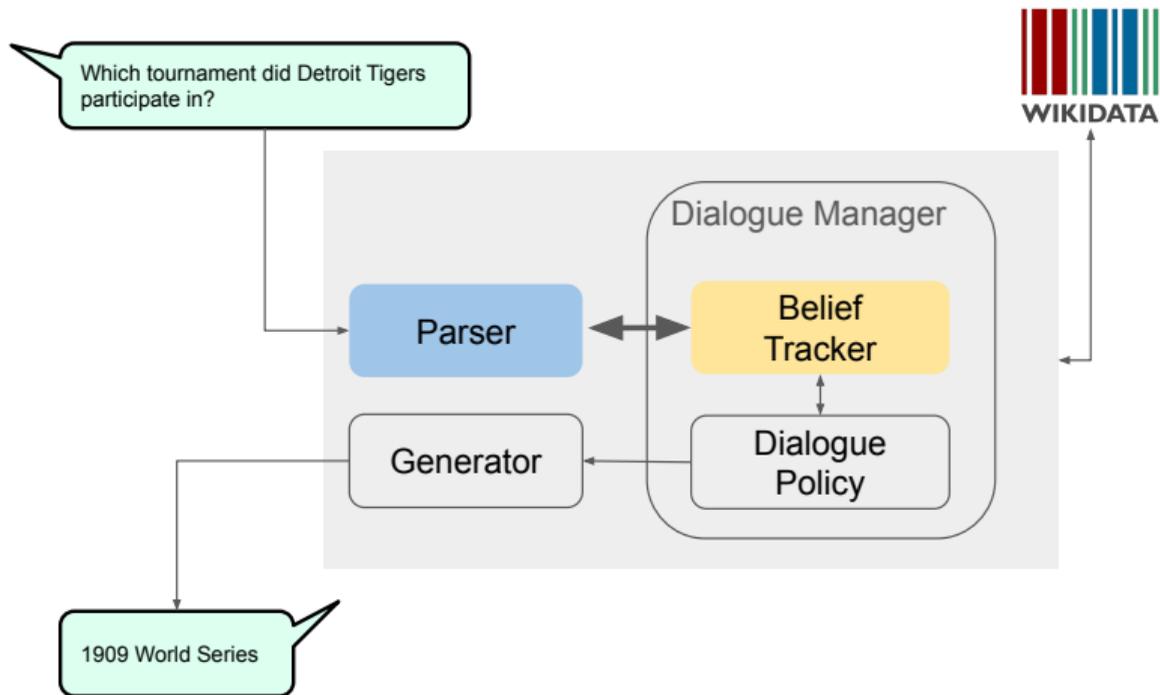


Conversational Question Answering over Knowledge Graphs



Semantic parsing: users interact with computers in human language!

Semantic Parsing for Conversational Question Answering over KGs



Grounding User Questions into SPARQL Queries over KGs

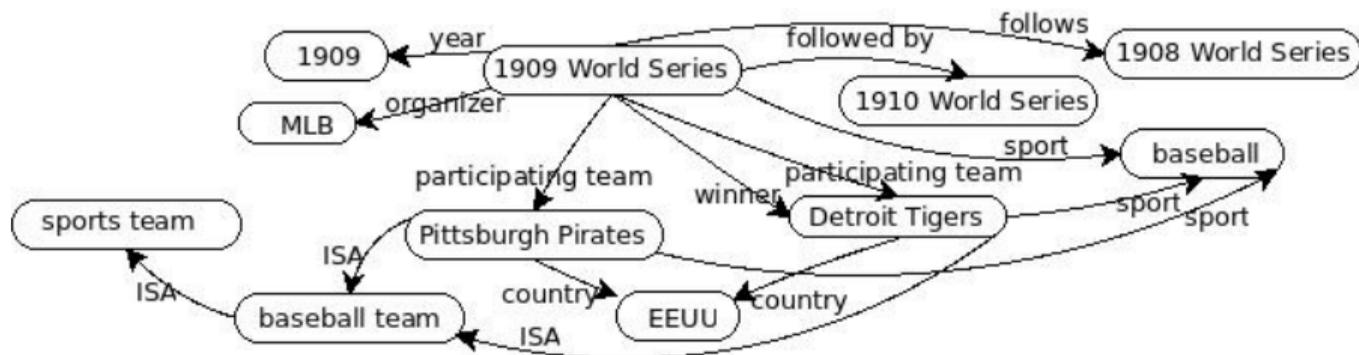
NL

Which sports team was the champion of that tournament?

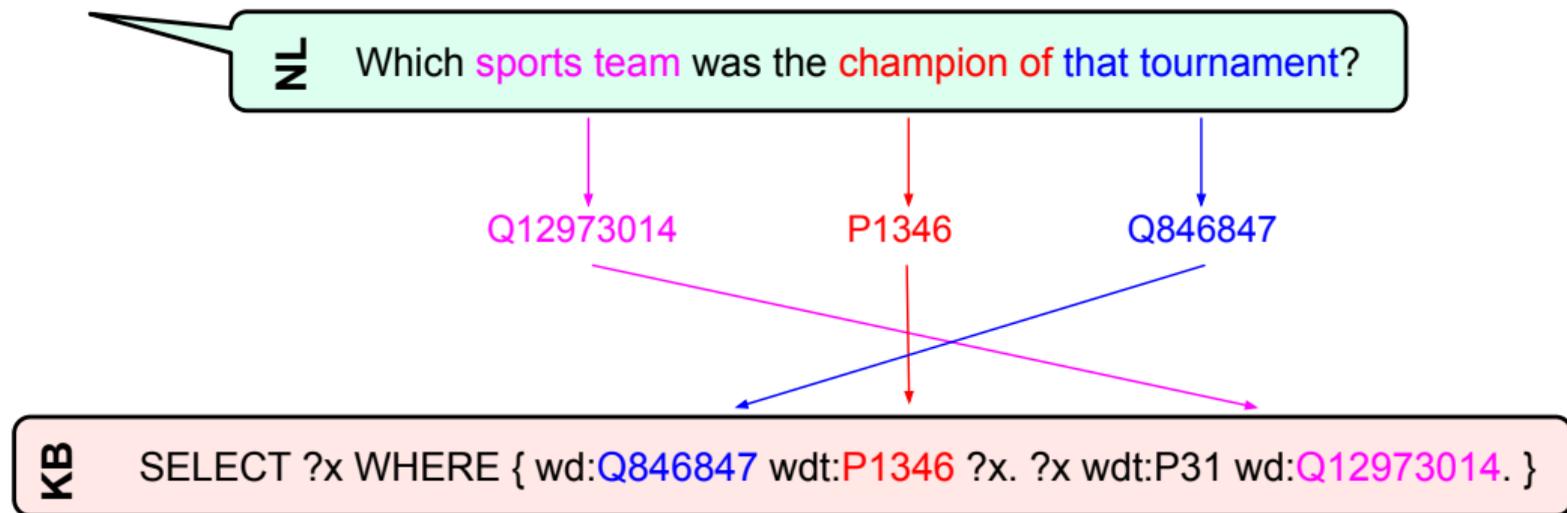
Parser

KB

```
SELECT ?x WHERE { wd:Q846847 wdt:P1346 ?x. ?x wdt:P31 wd:Q12973014. }
```



Grounding User Questions into SPARQL Queries over KGs

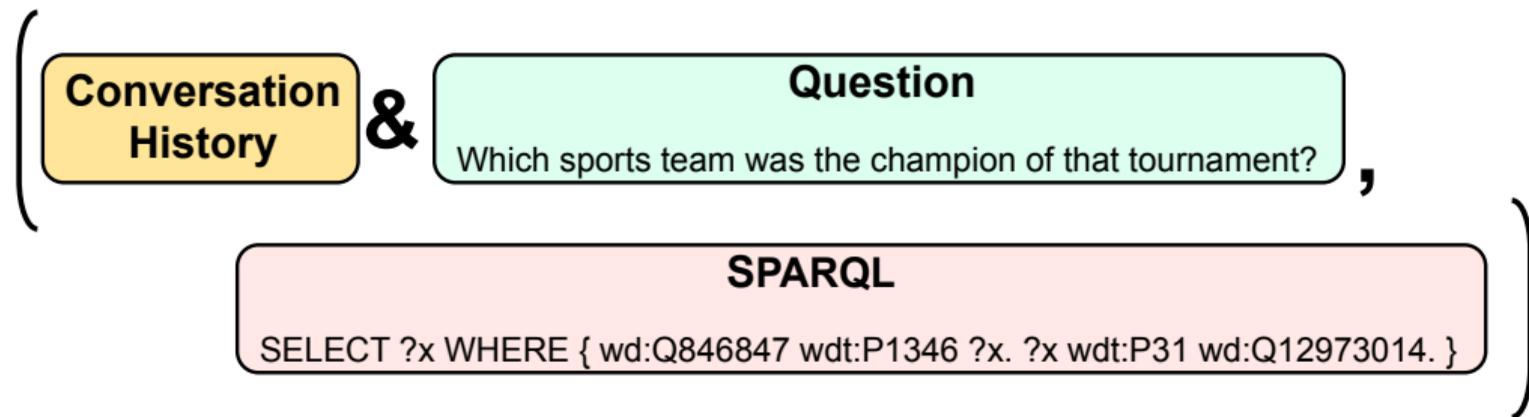


Challenges:

- Entity Linking & Scale (thousands of types and relations, millions of entities)
- Conversation phenomena (ellipsis, coreference, clarifications, topic shift)

Learning a Semantic Parser for Conversational Question Answering

Induce semantic parsers from (conversation history & question, SPARQL) pairs.



Existing Datasets

	Open Domain	Conversational	Executable
ATIS, CoSQL, SParC [Suhr et al. (2018); Yu et al. (2019b,a)]		✓	✓
LC-QuAD2.0, ComplexWebQuestions [Dubey et al. (2019); Talmor and Berant (2018)]	✓		✓
CSQA, ConvQuestions [Saha et al. (2018); Christmann et al. (2019, 2022)]	✓	✓	

Existing Datasets

	Open Domain	Conversational	Executable
ATIS, CoSQL, SParC [Suhr et al. (2018); Yu et al. (2019b,a)]		✓	✓
LC-QuAD2.0, ComplexWebQuestions [Dubey et al. (2019); Talmor and Berant (2018)]	✓		✓
CSQA, ConvQuestions [Saha et al. (2018); Christmann et al. (2019, 2022)]	✓	✓	
SPICE	✓	✓	✓

The SPICE Dataset

Nb. conversations	197k
Avg. turn length	9.5
Avg. entities per conversation	7.6
Avg. types per conversations	6.5
Nb. question intents	47

The SPICE Dataset

Nb. conversations	197k
Avg. turn length	9.5
Avg. entities per conversation	7.6
Avg. types per conversations	6.5
Nb. question intents	47

Simple questions

Which tournament did Detroit Tigers participate in?

The SPICE Dataset

Nb. conversations	197k
Avg. turn length	9.5
Avg. entities per conversation	7.6
Avg. types per conversations	6.5
Nb. question intents	47

Simple questions

Which tournament did Detroit Tigers participate in?

Reasoning questions

Which tournaments have less number of participating sports teams than 1909 World Series?

The SPICE Dataset

Nb. conversations	197k
Avg. turn length	9.5
Avg. entities per conversation	7.6
Avg. types per conversations	6.5
Nb. question intents	47

Simple questions

Which tournament did Detroit Tigers participate in?

Reasoning questions

Which tournaments have less number of participating sports teams than 1909 World Series?

Conversation

What are the countries of those sports teams? (co-reference)

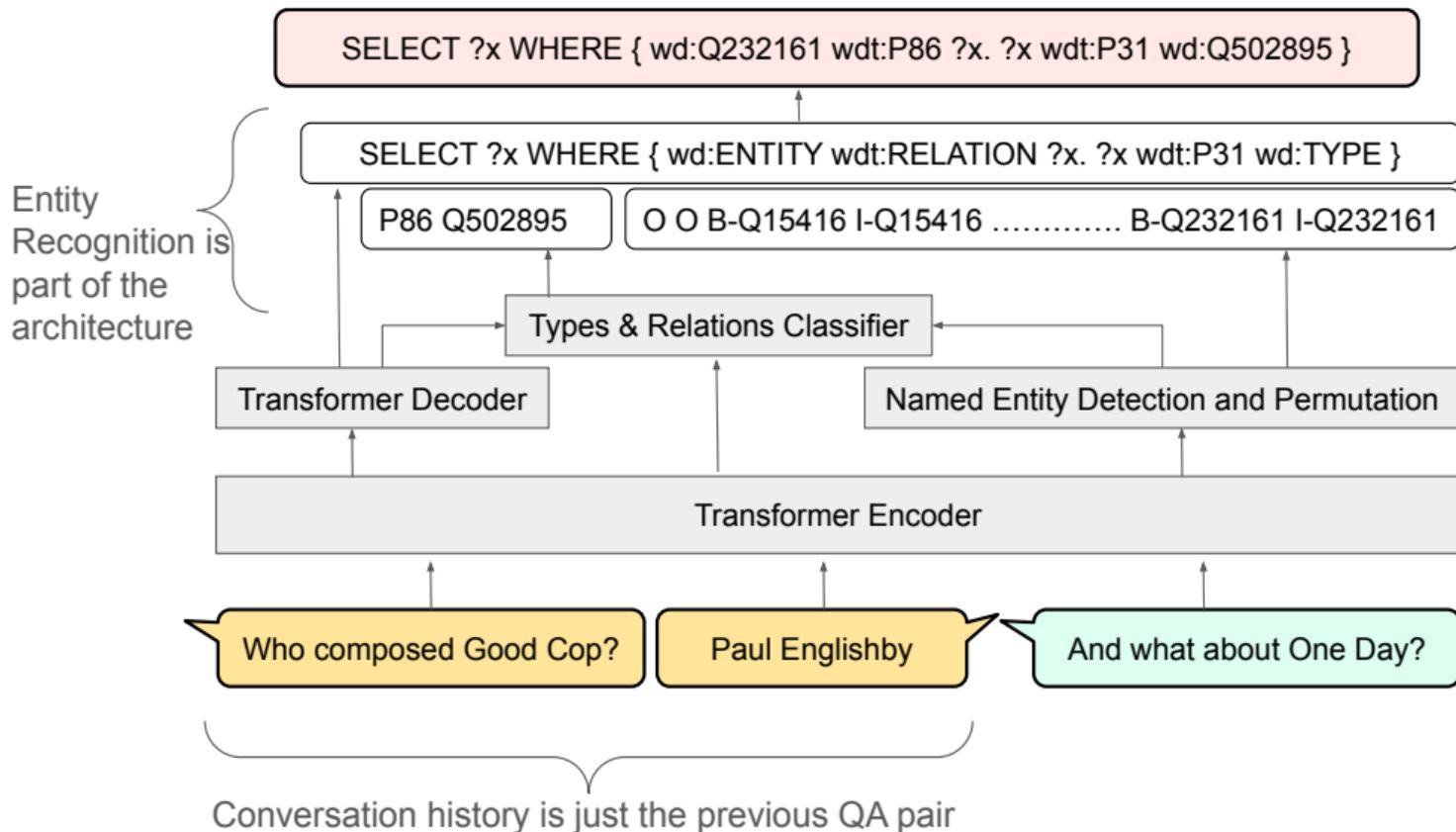
And what about 1910 World Series? (ellipsis)

How many sports teams participated in that tournament? (co-reference)

Did you mean Speed skating at the 2010 Winter Olympics? (clarification)

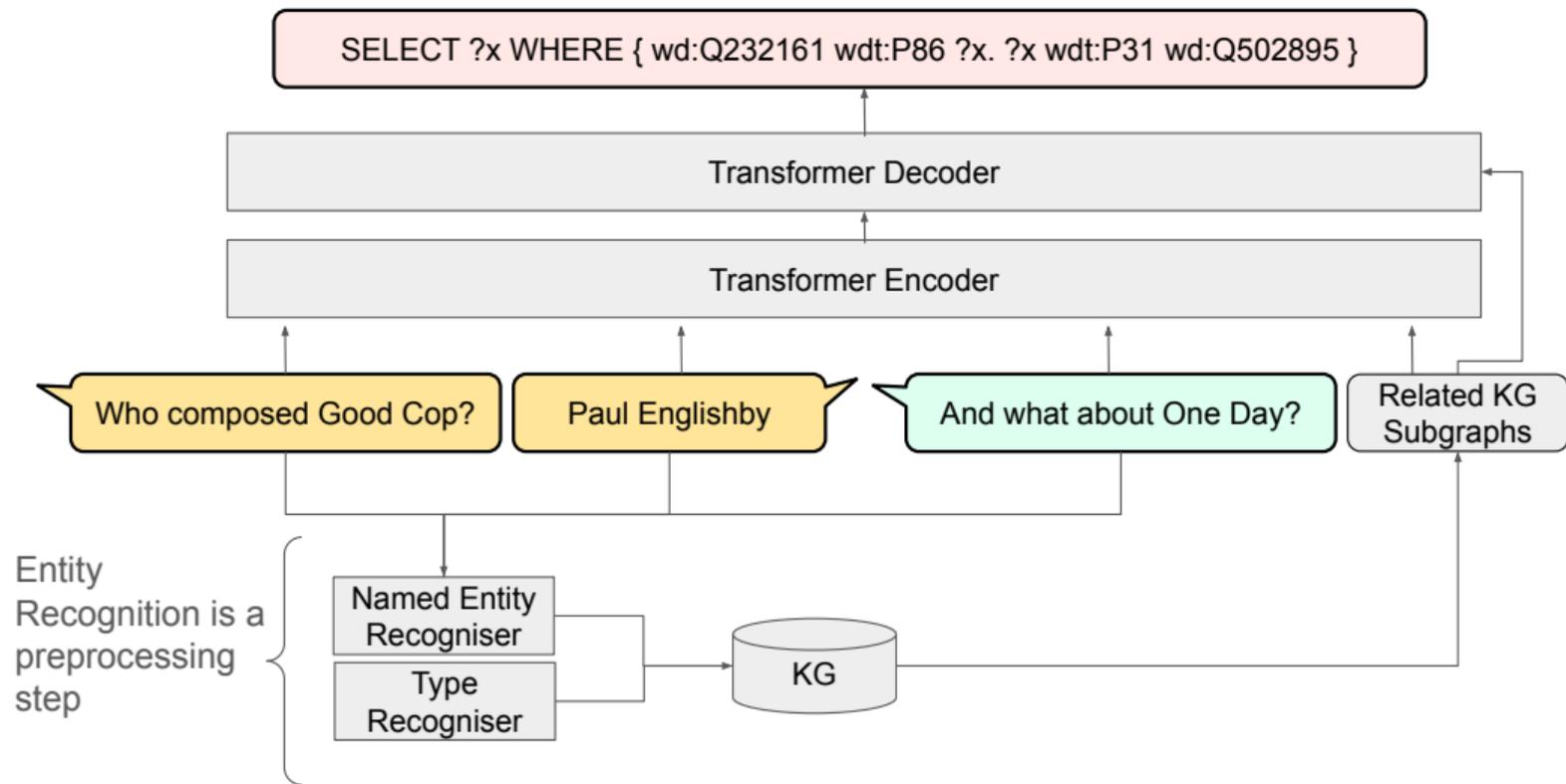
SPARQL Templates and Multiple Classifiers – Lasagne

[Kacupaj et al. (2021)]



Sequence-to-sequence with Dynamic Vocabularies – BertSP

[Gu et al. (2021)]

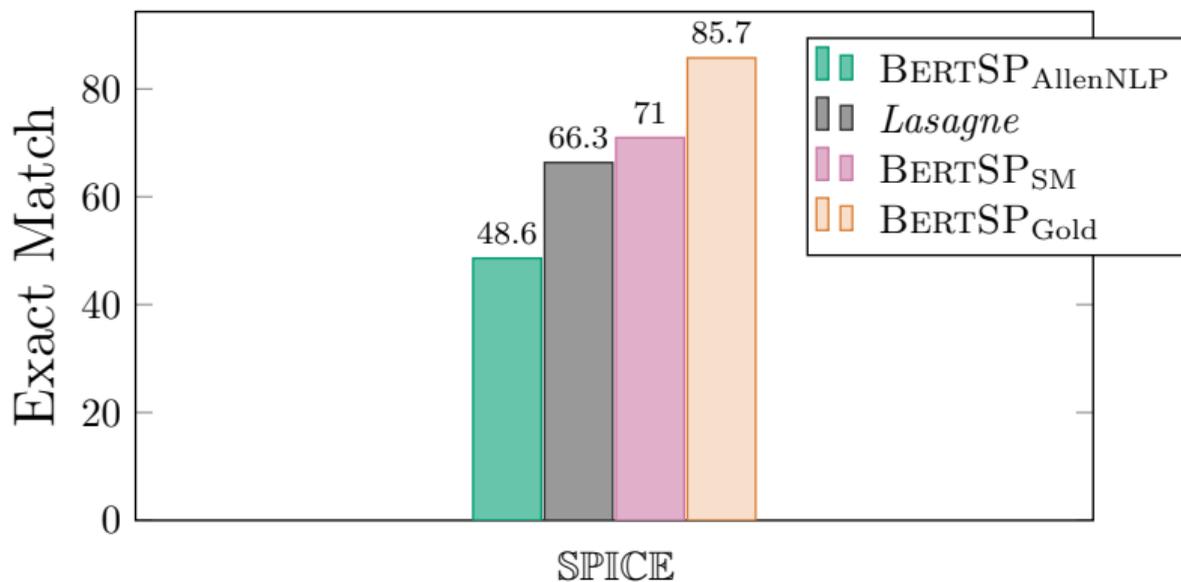


Average Baseline Performance Over Question Types

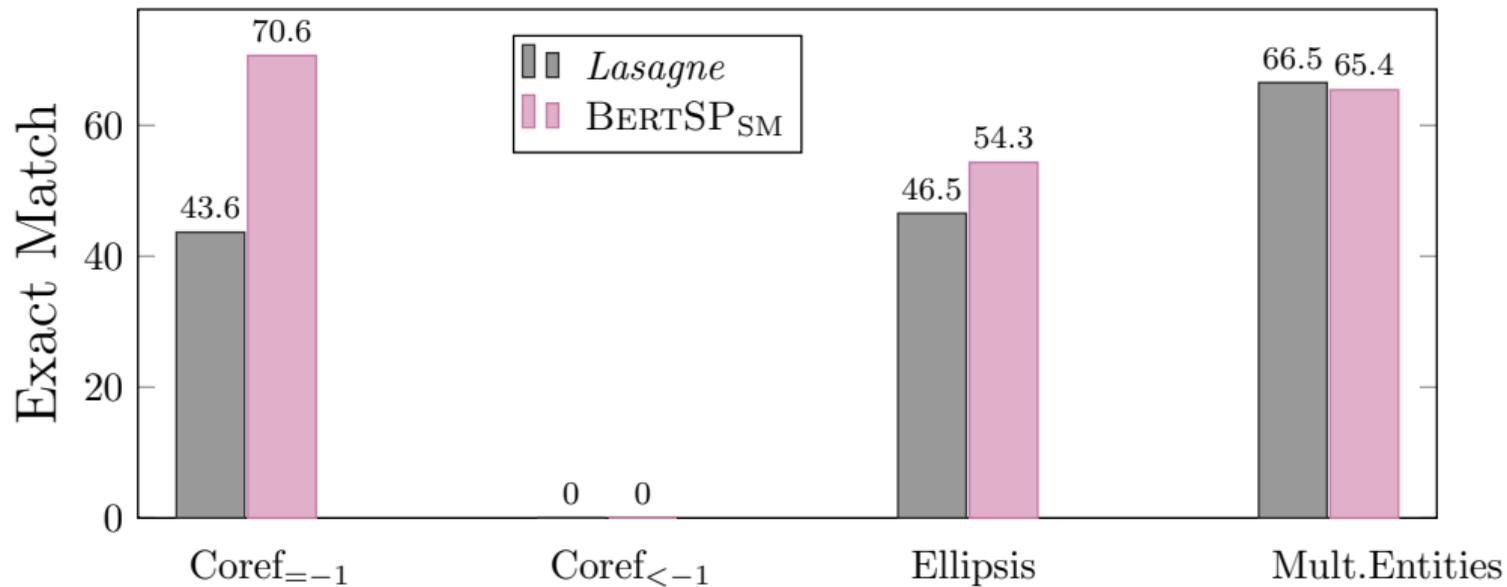
Exact Match Accuracy: proportion of input queries that are correctly parsed to their gold standard logical forms.

Average Baseline Performance Over Question Types

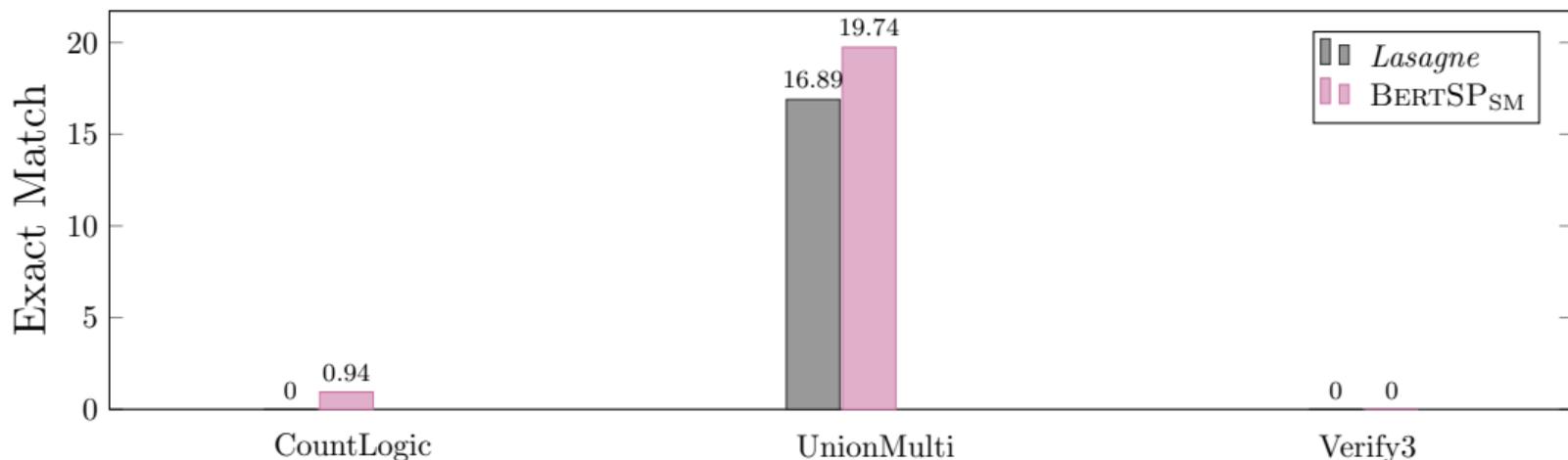
Exact Match Accuracy: proportion of input queries that are correctly parsed to their gold standard logical forms.



Models Struggle on Specific Linguistic Phenomena



And to Generalise to Unseen Questions



- **CountLogic**: a count operation over a union operator
- **UnionMulti**: a union operator over two graph patterns with different relations
- **Verify3**: verification questions with three entities

(Verify2) Is Zugspitze located in Germany?

(Verify3) Is Aix-en-Provence partner town of Baton Rouge and Hemmatabad, Alborz?

Conclusions

Directions to improve semantic parsing for conversational QA over KGs:

- improve entity recognition and linking
- better modelling of conversation history
- models with better generalisation capabilities

Code and SPICE @ <https://github.com/EdinburghNLP/SPICE>

References I

- I. Androutsopoulos, G.D. Ritchie, and P. Thanisch. 1995. Natural language interfaces to databases: an introduction. Natural Language Engineering, 1(1):2981.
- Philipp Christmann, Rishiraj Saha Roy, Abdalghani Abujabal, Jyotsna Singh, and Gerhard Weikum. 2019. Look before you hop: Conversational question answering over knowledge graphs using judicious context expansion. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pages 729–738, Beijing, China. ACM.
- Philipp Christmann, Rishiraj Saha Roy, and Gerhard Weikum. 2022. Conversational question answering on heterogeneous sources. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 144–154, Madrid, Spain. ACM.
- Mohnish Dubey, Debayan Banerjee, Abdelrahman Abdelkawi, and Jens Lehmann. 2019. Lc-quad 2.0: A large dataset for complex question answering over wikidata and dbpedia. In The Semantic Web – ISWC 2019, pages 69–78, Cham. Springer International Publishing.
- Jianfeng Gao, Chenyan Xiong, Paul Bennett, and Nick Craswell. 2023. Neural Approaches to Conversational Information Retrieval. The Information Retrieval series. Springer.
- Yu Gu, Sue Kase, Michelle Vanni, Brian Sadler, Percy Liang, Xifeng Yan, and Yu Su. 2021. Beyond i.i.d.: Three levels of generalization for question answering on knowledge bases. In Proceedings of the Web Conference 2021, WWW '21, page 34773488, New York, NY, USA. Association for Computing Machinery.
- Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. 2017. Search-based neural structured learning for sequential question answering. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1821–1831, Vancouver, Canada. Association for Computational Linguistics.
- Endri Kacupaj, Joan Plepi, Kuldeep Singh, Harsh Thakkar, Jens Lehmann, and Maria Maleshkova. 2021. Conversational question answering over knowledge graphs with transformer and graph attention networks. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 850–862, Online. Association for Computational Linguistics.
- Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR '17, page 117126, New York, NY, USA. Association for Computing Machinery.
- Amrita Saha, Vardaan Pahuja, Mitesh Khapra, Karthik Sankaranarayanan, and Sarath Chandar. 2018. Complex sequential question answering: Towards learning to converse over linked question answer pairs with a knowledge graph. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), pages 705–713, New Orleans, Louisiana, USA. AAAI Press.
- Alane Suhr, Srinivasan Iyer, and Yoav Artzi. 2018. Learning to map context-dependent sentences to executable formal queries. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 2238–2249, New Orleans, Louisiana. Association for Computational Linguistics.

References II

- Alon Talmor and Jonathan Berant. 2018. The web as a knowledge-base for answering complex questions. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 641–651, New Orleans, Louisiana. Association for Computational Linguistics.
- Tao Yu, Rui Zhang, Heyang Er, Suyi Li, Eric Xue, Bo Pang, Xi Victoria Lin, Yi Chern Tan, Tianze Shi, Zihan Li, Youxuan Jiang, Michihiro Yasunaga, Sungrok Shim, Tao Chen, Alexander Fabbri, Zifan Li, Luyao Chen, Yuwen Zhang, Shreya Dixit, Vincent Zhang, Caiming Xiong, Richard Socher, Walter Lasecki, and Dragomir Radev. 2019a. CoSQL: A conversational text-to-SQL challenge towards cross-domain natural language interfaces to databases. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 1962–1979, Hong Kong, China. Association for Computational Linguistics.
- Tao Yu, Rui Zhang, Michihiro Yasunaga, Yi Chern Tan, Xi Victoria Lin, Suyi Li, Heyang Er, Irene Li, Bo Pang, Tao Chen, Emily Ji, Shreya Dixit, David Proctor, Sungrok Shim, Jonathan Kraft, Vincent Zhang, Caiming Xiong, Richard Socher, and Dragomir Radev. 2019b. SPaC: Cross-domain semantic parsing in context. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4511–4523, Florence, Italy. Association for Computational Linguistics.