Certainly! Here's a SPARQL query that you can use to retrieve the list of countries in the European Union (EU) from DBPedia:

sparql

Copy code

```sparql
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>

SELECT ?country ?countryName WHERE {
  ?country dbpedia-owl:membership <http://dbpedia.org/resource/European_Unio
  ?country dbp:name ?countryName .
  ?country a dbpedia-owl:Country .
}
ORDER BY ?countryName
```

is dbp:membership of	• dbr:Paris_Club • dbr:Global_Health_Security_Initiative • dbr:European_Economic_Area • dbr:European_Go_Federation

dcterms:subject	• dbc:Italy • dbc:G20_nations • dbc:Southwestern_European_countries • dbc:Italian-speaking_countries_and_territories • dbc:Transcontinental_countries • dbc:1861_establishments_in_Europe • dbc:States_and_territories_established_in_1861 • dbc:Member_states_of_the_United_Nations • dbc:Countries_in_Europe • dbc:Member_states_of_NATO • dbc:Member_states_of_the_Council_of_Europe • dbc:Member_states_of_the_Union_for_the_Mediterranean • dbc:Republics • dbc:Southern_European_countries • dbc:Member_states_of_the_European_Union • dbc:G7_nations

Default Data Set Name (Graph IRI)

http://dbpedia.org

Query Text

```
PREFIX dbo: <http://dbpedia.org/ontology/>
PREFIX dbp: <http://dbpedia.org/property/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dbc: <http://dbpedia.org/resource/Category>

SELECT ?country ?countryName WHERE {
  ?country dcterms:subject <http://dbpedia.org/resource/Category:Member_states_of_the_European_Union> .
  ?country dbp:commonName ?countryName .
  ?country a dbo:Country .
} ORDER BY ?countryName
```

Results Format HTML

Execute Query Reset

SPARQL | HTML5 table

country	countryName
http://dbpedia.org/resource/Austria	"Austria"@en
http://dbpedia.org/resource/Belgium	"Belgium"@en
http://dbpedia.org/resource/Bulgaria	"Bulgaria"@en
http://dbpedia.org/resource/Croatia	"Croatia"@en
http://dbpedia.org/resource/Cyprus	"Cyprus"@en
http://dbpedia.org/resource/Estonia	"Estonia"@en
http://dbpedia.org/resource/Finland	"Finland"@en
http://dbpedia.org/resource/France	"France"@en
http://dbpedia.org/resource/Germany	"Germany"@en

Asking Questions in Natural Language

Pasquale Minervini
p.minervini@ed.ac.uk

Main References

1. ACL 2023 Tutorial: Retrieval-based Language Models and Applications — https://acl2023-retrieval-lm.github.io/
2. Recent Advances in Text-to-SQL: A Survey of What We Have and What We Expect (COLING 2022) — https://aclanthology.org/2022.coling-1.190.pdf

Did Socrates use a laptop?

Geva et al. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies

Did Socrates use a laptop?

Q1 Did Aristotle use a laptop?

Geva et al. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies

Did Socrates use a laptop?

Q1 Did Aristotle use a laptop?

Q2 Was Aristotle alive when the laptop was invented?

Geva et al. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies

Did Socrates use a laptop?

Q1 Did Aristotle use a laptop?

Q2 Was Aristotle alive when the laptop was invented?

implicit

explicit

D

1. When did Aristotle live?

2. When was the laptop invented?

3. Is #2 before #1?

Geva et al. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies

Did Socrates use a laptop?

Q1 Did Aristotle use a laptop?

Q2 Was Aristotle alive when the laptop was invented?

implicit

explicit

D

1. When did Aristotle live?
2. When was the laptop invented?
3. Is #2 before #1?

E

"Aristotle (384 –322 BC) was a philosopher…"

"The first laptop was… in 1980."

Geva et al. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies

Did Socrates use a laptop?

Q1 Did Aristotle use a laptop?

Q2 Was Aristotle alive when the laptop was invented?

implicit

explicit

D

1. When did Aristotle live?
2. When was the laptop invented?
3. Is #2 before #1?

A No

E

"Aristotle (384 –322 BC) was a philosopher…"

"The first laptop was… in 1980."

Geva et al. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies

Did Socrates use a laptop?

No, Socrates did not use a laptop. Socrates lived in ancient Greece from 469/470 BCE to 399 BCE, long before the invention of laptops. Laptops were not developed until the 1970s, more than 2000 years after Socrates' death. Instead of using a laptop, Socrates engaged in philosophical conversations with his students and peers, often in public places like the agora, the marketplace of ancient Athens. His ideas were later recorded in the writings of his students, particularly Plato, who documented many of Socrates' dialogues in his works.

What happened between Ukraine and Russia on February 2022?

I apologize, but as of my last knowledge update in September 2021, I do not have information on events that occurred in February 2022 or beyond. It's important to stay updated with reliable news sources or consult recent information to learn about the specific events that took place between Ukraine and Russia during that time.

What happened to Mahsa Amini?

As of my last knowledge update in September 2021, I do not have information on Mahsa Amini. It is possible that this person became known or was involved in an event that occurred after my knowledge cutoff. To obtain the most accurate and up-to-date information on Mahsa Amini, I recommend referring to reliable news sources or conducting an online search for recent developments.

When did OpenAI launch GPT-4?

As of my last knowledge update in September 2021, OpenAI had not released GPT-4. At that time, the most recent version available was GPT-3, which was released in June 2020. However, please note that developments may have occurred since then, and it is recommended to refer to the latest news and updates from OpenAI or reliable sources to obtain information on any potential release of GPT-4 or subsequent versions.

Large Language Models

Extremely impressive:

Large Language Models

Extremely impressive:

- Store large quantities of knowledge in their parameters ✅

Large Language Models

Extremely impressive:

- Store large quantities of knowledge in their parameters ✅
- Impressive results on a variety of tasks ✅

Large Language Models

Extremely impressive:

- Store large quantities of knowledge in their parameters ✅
- Impressive results on a variety of tasks ✅
- Widely applicable ✅

However ..

Large Language Models

Extremely impressive:

- Store large quantities of knowledge in their parameters ✅
- Impressive results on a variety of tasks ✅
- Widely applicable ✅

However ..

- Difficult to update and control knowledge ❌

Large Language Models

Extremely impressive:

- Store large quantities of knowledge in their parameters ✅
- Impressive results on a variety of tasks ✅
- Widely applicable ✅

However ..

- Difficult to update and control knowledge ❌
- Black-boxes — lack of provenance and interpretability ❌

Large Language Models

Extremely impressive:

- Store large quantities of knowledge in their parameters ✅
- Impressive results on a variety of tasks ✅
- Widely applicable ✅

However ..

- Difficult to update and control knowledge ❌
- Black-boxes — lack of provenance and interpretability ❌
- Very large and expensive ❌

Integrating Factual Knowledge in Neural Models

(via retrieval-augmentation and differentiable external memories)

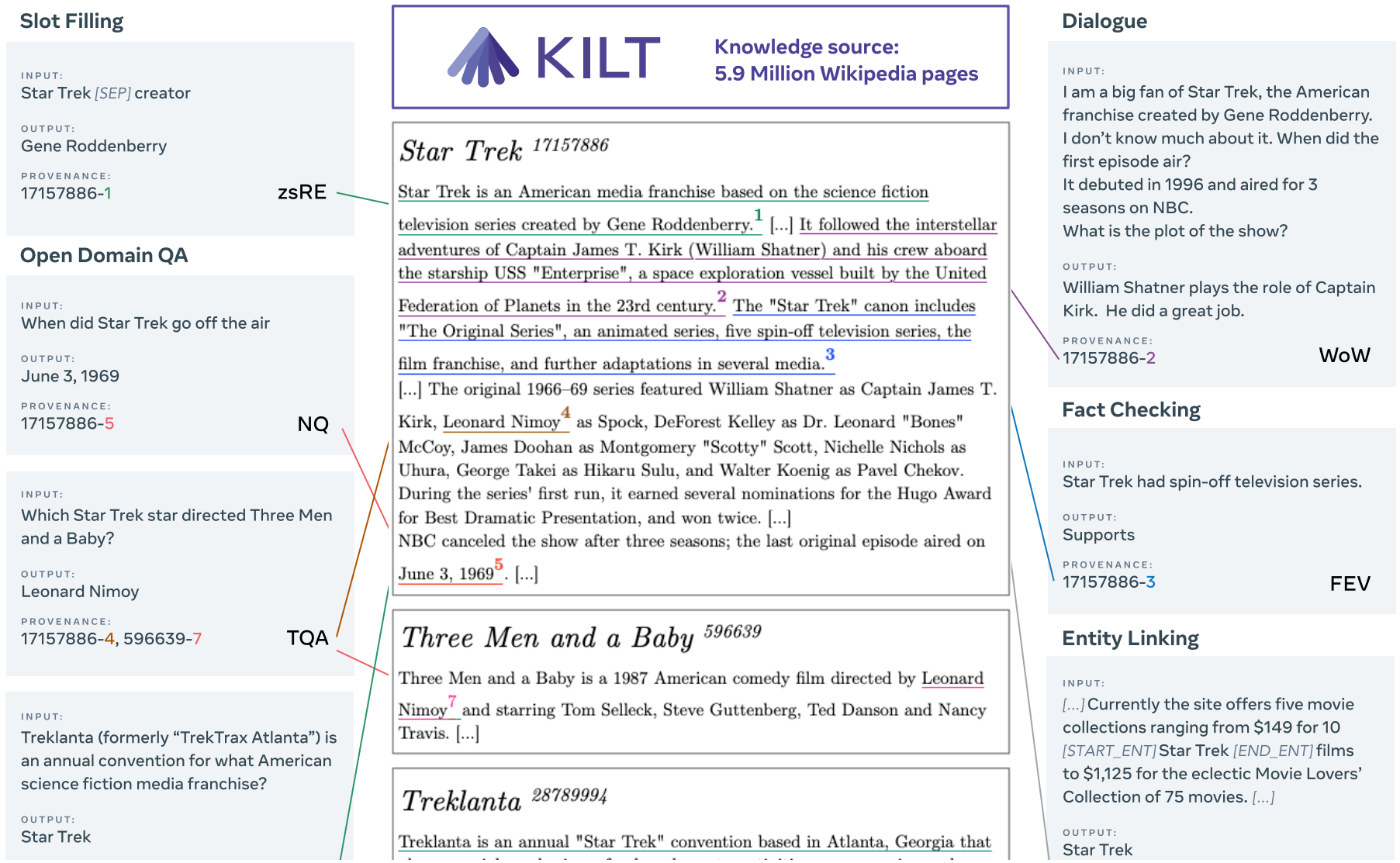
Knowledge-Intensive NLP Tasks

question (string)	answer (sequence)
"where did they film hot tub time machine"	["Fernie Alpine Resort"]
"who has the right of way in international waters"	["Neither vessel"]
"who does annie work for attack on titan"	["Marley"]
"when was the immigration reform and control act passed"	["November 6, 1986"]
"when was puerto rico added to the usa"	["1950"]
"who has been chosen for best supporting actress in 64 national filmfare award"	["Zaira Wasim"]
"which side of the white house is the front"	["North"]
"names of the metropolitan municipalities in south africa"	["Mangaung Metropolitan Municipality", "Nelson Mandela Bay Metropolitan Municipality", "eThekwini…

Knowledge-Intensive NLP Tasks

input (string)	output (list)
"Nikolaj Coster-Waldau worked with the Fox Broadcasting Company."	[{ "answer": "SUPPORTS", "provenance": [] }]
"Roman Atwood is a content creator."	[{ "answer": "SUPPORTS", "provenance": [{ "bleu_sc
"History of art includes architecture, dance, sculpture, music, painting, poetry literature,…	[{ "answer": "SUPPORTS", "provenance": [] }]
"Adrienne Bailon is an accountant."	[{ "answer": "REFUTES", " "provenance": [{ "bleu_sc
"Homeland is an American television spy thriller based on the Israeli television series Prisoners…	[{ "answer": "SUPPORTS", "provenance": [{ "bleu_sc

The KILT Benchmark

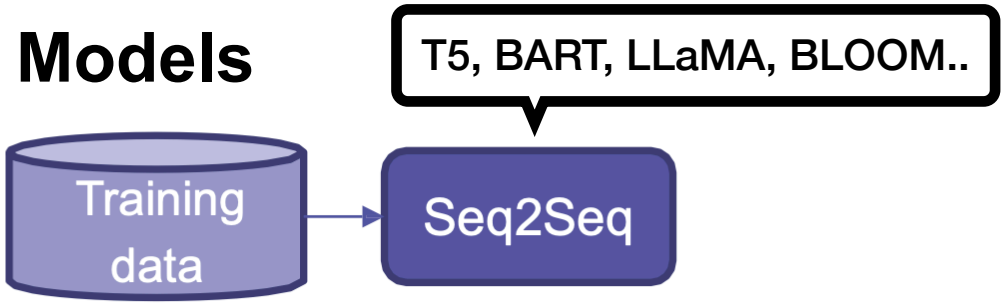


- Open-Domain Question Answering (Natural Questions, TriviaQA, HotPotQA, ELI5)
- Fact-Checking (FEVER)
- Slot Filling (T-REx, zsRE)
- Dialogue (Wizard of Wikipedia)
- Entity Linking (AIDA, WNED-WIKI, WNED-CWEB)

Petroni et al. KILT: A Benchmark for Knowledge Intensive Language Tasks

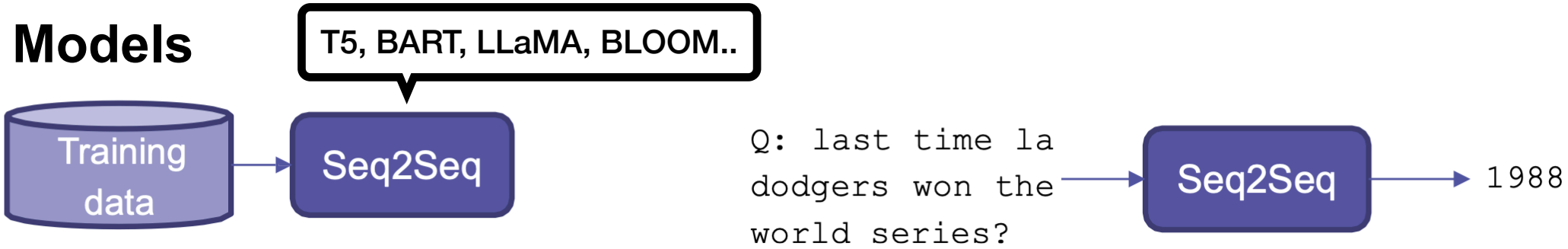
Solving Knowledge-Intensive NLP Tasks

Parametric Models



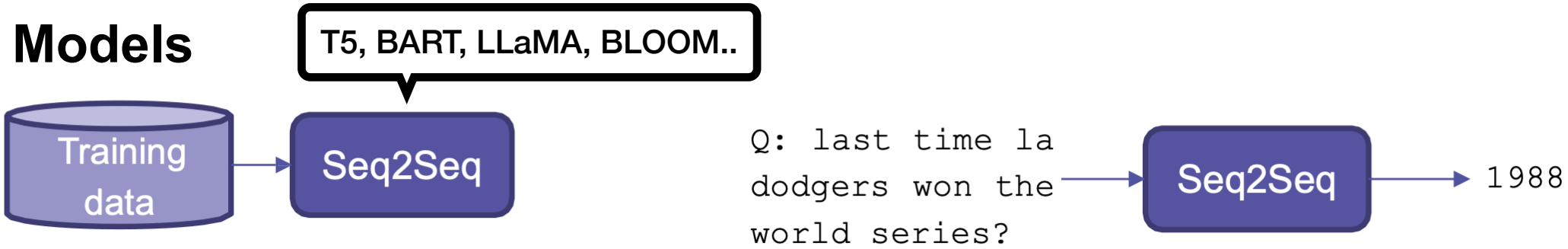
Solving Knowledge-Intensive NLP Tasks

Parametric Models



Solving Knowledge-Intensive NLP Tasks

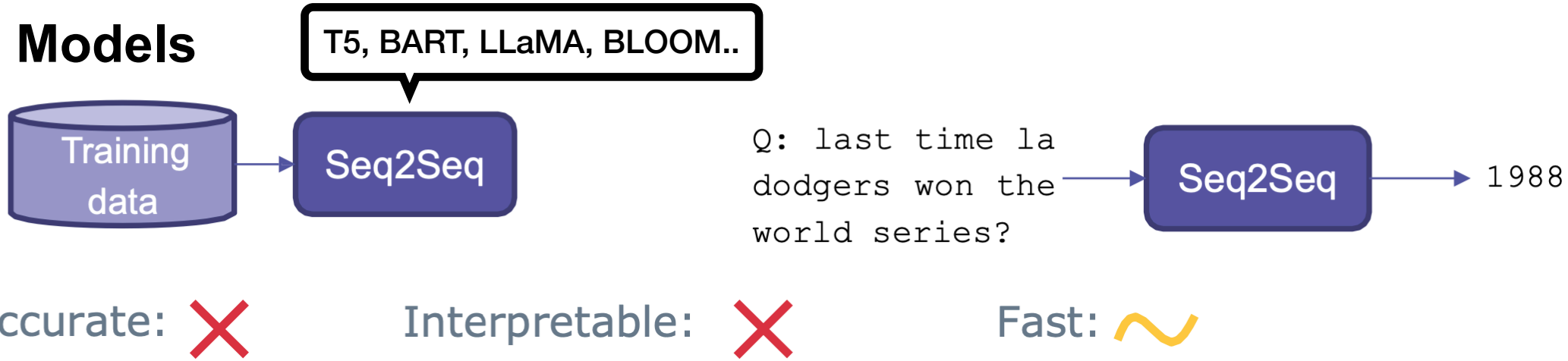
Parametric Models



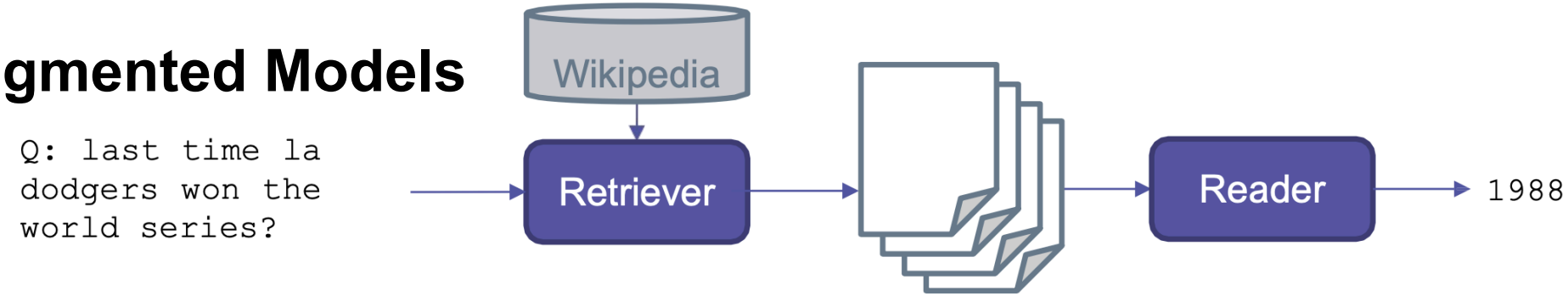
Accurate: ✗ Interpretable: ✗ Fast: ~

Solving Knowledge-Intensive NLP Tasks

Parametric Models

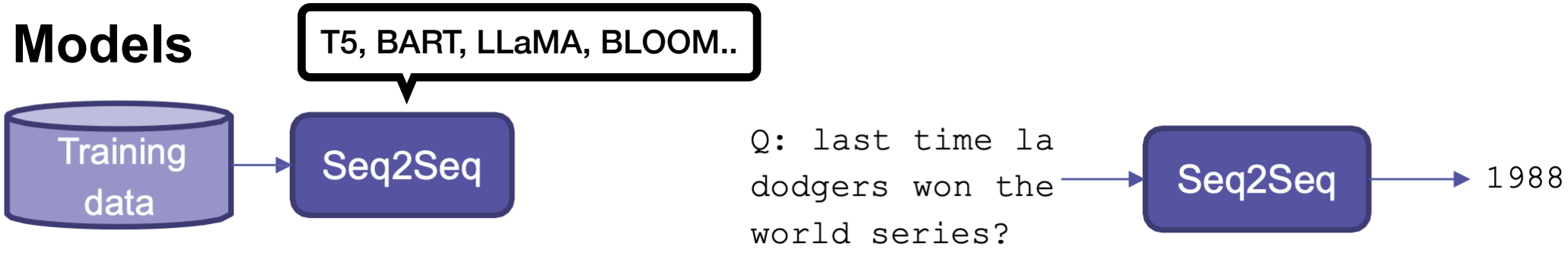


Retrieval-Augmented Models



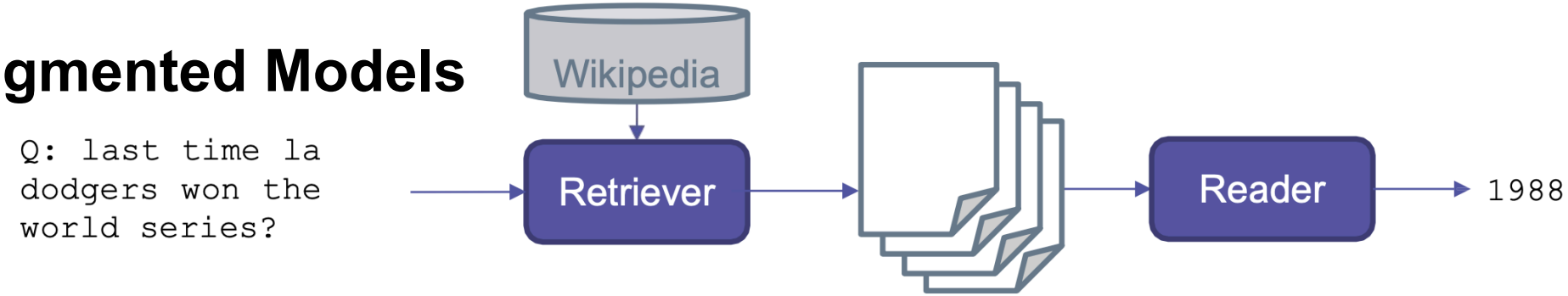
Solving Knowledge-Intensive NLP Tasks

Parametric Models



Accurate: ✗ Interpretable: ✗ Fast: ~

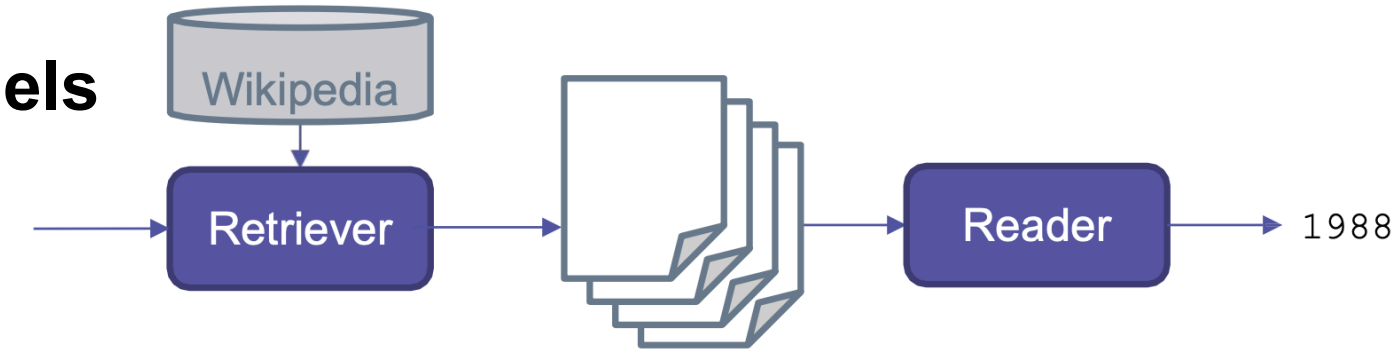
Retrieval-Augmented Models



Accurate: ✓ Interpretable: ~ Fast: ✗

Solving Knowledge-Intensive NLP Tasks

Retrieval-Augmented Models

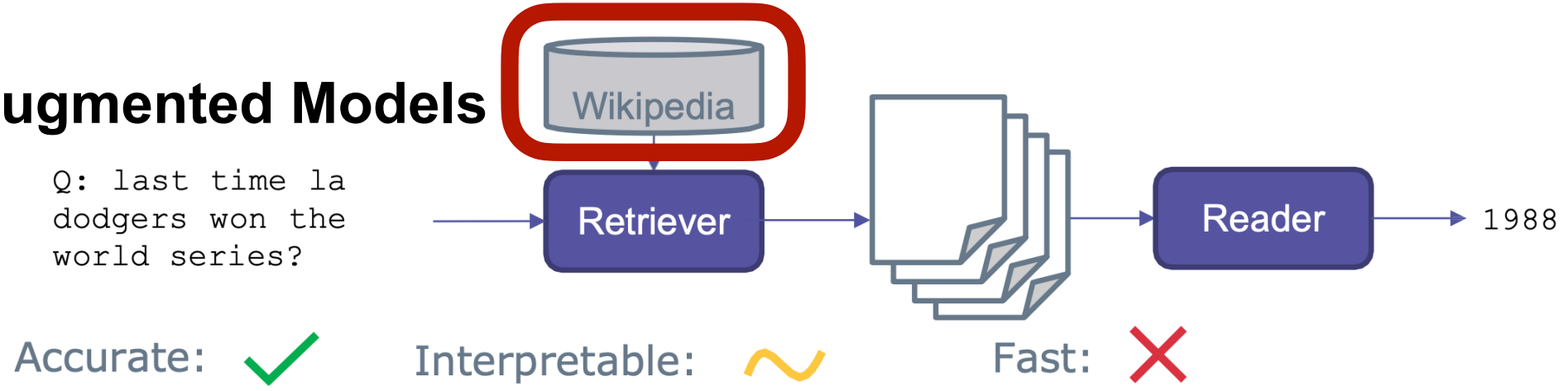


Accurate: ✓ Interpretable: ~ Fast: ✗

Retrieval-Augmented Models in a nutshell:

Solving Knowledge-Intensive NLP Tasks

Retrieval-Augmented Models

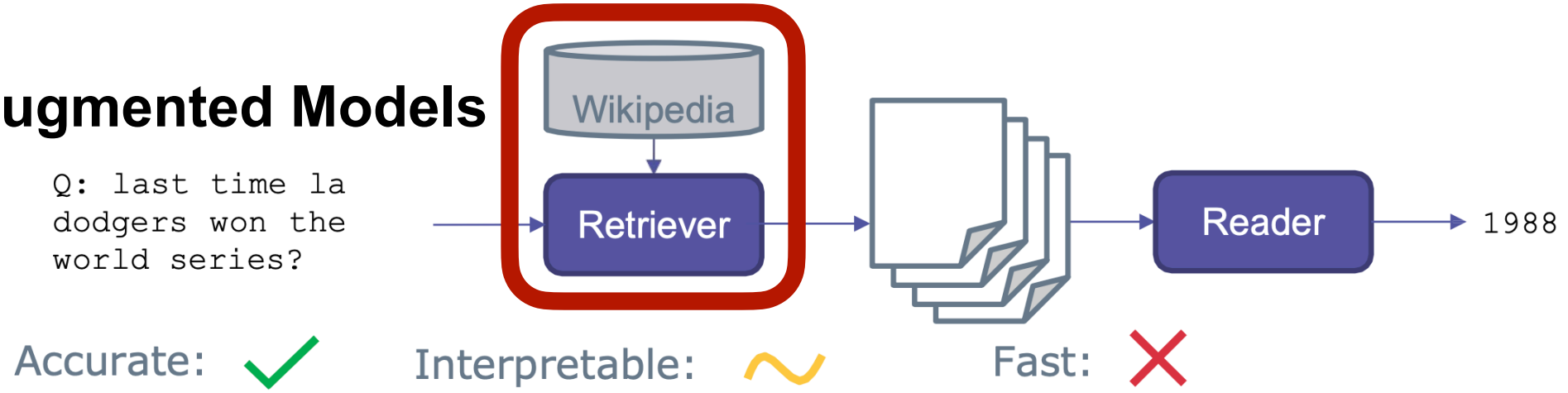


Retrieval-Augmented Models in a nutshell:

- Document representations are stored in a **large (1M-1B entries) dense vector index**

Solving Knowledge-Intensive NLP Tasks

Retrieval-Augmented Models

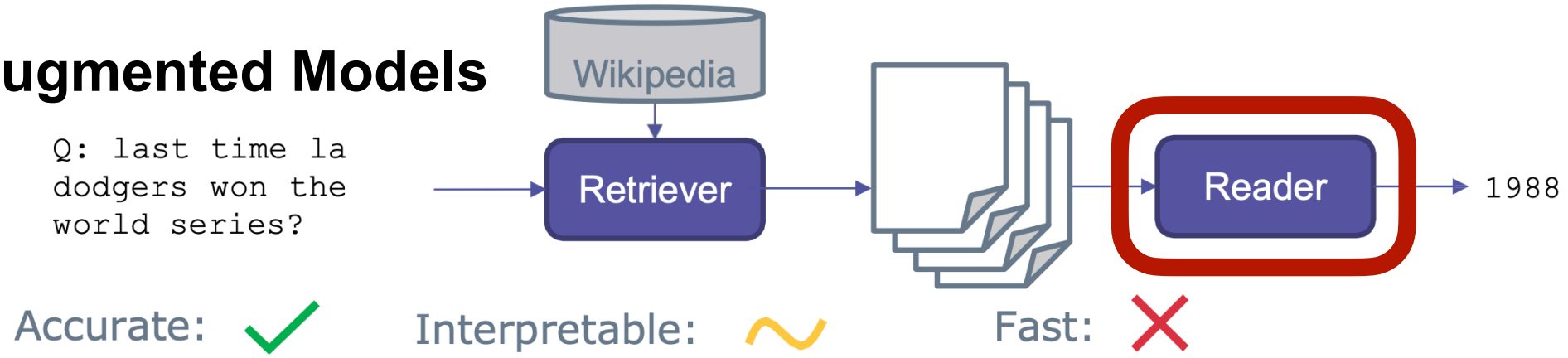


Retrieval-Augmented Models in a nutshell:

- Document representations are stored in a **large (1M-1B entries) dense vector index**
- A **retriever** retrieves the most relevant document are retrieved via **similarity search**

Solving Knowledge-Intensive NLP Tasks

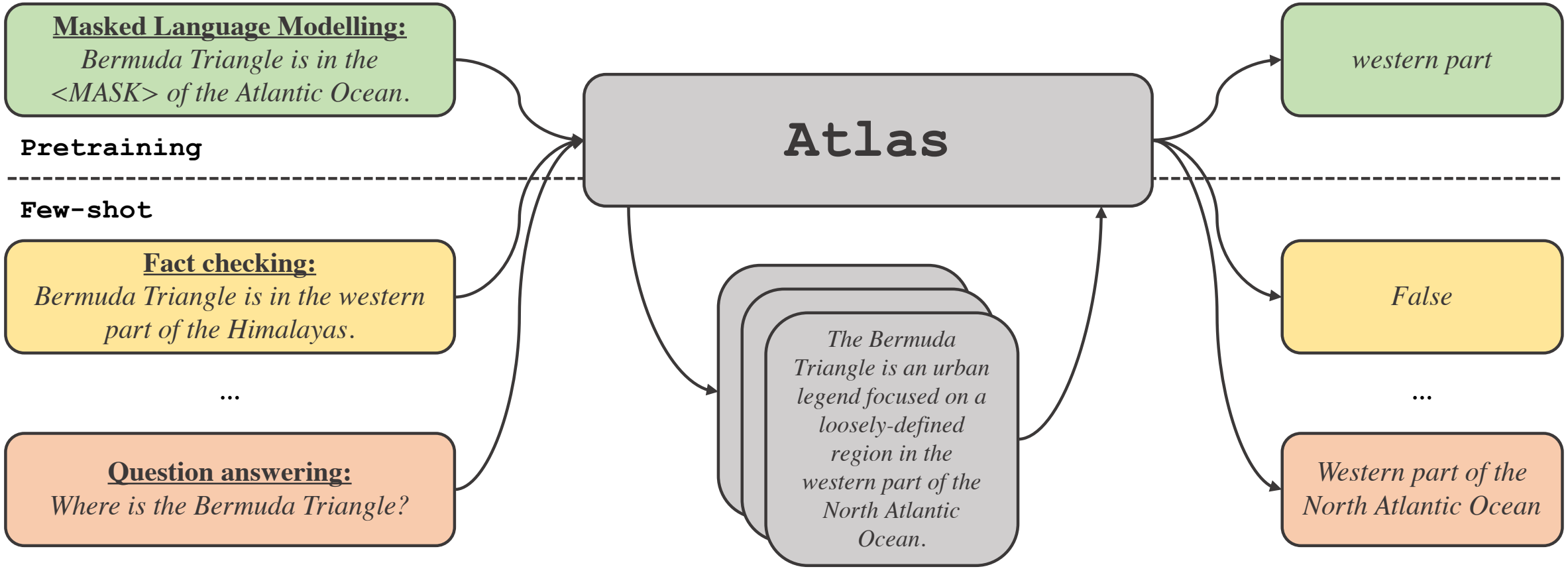
Retrieval-Augmented Models



Retrieval-Augmented Models in a nutshell:

- Document representations are stored in a **large (1M-1B entries) dense vector index**
- A **retriever** retrieves the most relevant document are retrieved via **similarity search**
- A **reader** produces the output conditioned on the input and the **retrieved documents**

ATLAS — A Retrieval-Augmented Model



Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

ATLAS — Open-Domain QA

Model	NQ		TriviaQA filtered		TriviaQA unfiltered	
	64-shot	Full	64-shot	Full	64-shot	Full
GPT-3 (Brown et al., 2020)	29.9	-	-	-	71.2	-
Gopher (Rae et al., 2021)	28.2	-	57.2	-	61.3	-
Chinchilla (Hoffmann et al., 2022)	35.5	-	64.6	-	72.3	-
PaLM (Chowdhery et al., 2022)	39.6	-	-	-	81.4	-
RETRO (Borgeaud et al., 2021)	-	45.5	-	-	-	-
FiD (Izacard & Grave, 2020)	-	51.4	-	67.6	-	80.1
FiD-KD (Izacard & Grave, 2021)	-	54.7	-	73.3	-	-
R2-D2 (Fajcik et al., 2021)	-	55.9	-	69.9	-	-

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

ATLAS — Open-Domain QA

Model	NQ		TriviaQA filtered		TriviaQA unfiltered	
	64-shot	Full	64-shot	Full	64-shot	Full
GPT-3 (Brown et al., 2020)	29.9	-	-	-	71.2	-
Gopher (Rae et al., 2021)	28.2	-	57.2	-	61.3	-
Chinchilla (Hoffmann et al., 2022)	35.5	-	64.6	-	72.3	-
PaLM (Chowdhery et al., 2022)	39.6	-	-	-	81.4	-
RETRO (Borgeaud et al., 2021)	-	45.5	-	-	-	-
FiD (Izacard & Grave, 2020)	-	51.4	-	67.6	-	80.1
FiD-KD (Izacard & Grave, 2021)	-	54.7	-	73.3	-	-
R2-D2 (Fajcik et al., 2021)	-	55.9	-	69.9	-	-

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

ATLAS — Open-Domain QA

Model	NQ		TriviaQA filtered		TriviaQA unfiltered	
	64-shot	Full	64-shot	Full	64-shot	Full
GPT-3 (Brown et al., 2020)	29.9	-	-	-	71.2	-
Gopher (Rae et al., 2021)	28.2	-	57.2	-	61.3	-
Chinchilla (Hoffmann et al., 2022)	35.5	-	64.6	-	72.3	-
PaLM (Chowdhery et al., 2022)	39.6	-	-	-	81.4	-
RETRO (Borgeaud et al., 2021)	-	45.5	-	-	-	-
FiD (Izacard & Grave, 2020)	-	51.4	-	67.6	-	80.1
FiD-KD (Izacard & Grave, 2021)	-	54.7	-	73.3	-	-
R2-D2 (Fajcik et al., 2021)	-	55.9	-	69.9	-	-
ATLAS	**42.4**	**60.4**	**74.5**	**79.8**	**84.7**	**89.4**

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

Generalisation

Model	Open Natural Questions				TriviaQA				WebQuestions			
	Total	Question Overlap	Answer Overlap Only	No Overlap	Total	Question Overlap	Answer Overlap Only	No Overlap	Total	Question Overlap	Answer Overlap Only	No Overlap

The question appears in the training set

The answer appears in the training set

Lewis et al. Overlap in Open-Domain Question Answering Datasets

Generalisation

Model		Open Natural Questions				TriviaQA				WebQuestions			
		Total	Question Overlap	Answer Overlap Only	No Overlap	Total	Question Overlap	Answer Overlap Only	No Overlap	Total	Question Overlap	Answer Overlap Only	No Overlap
Closed book	T5-11B+SSM	36.6	77.2	22.2	9.4	-	-	-	-	44.7	82.1	44.5	22.0
	BART	26.5	67.6	10.2	0.8	26.7	67.3	16.3	0.8	27.4	71.5	20.7	1.6

The question appears in the training set

The answer appears in the training set

Lewis et al. Overlap in Open-Domain Question Answering Datasets

Generalisation

Model		Open Natural Questions				TriviaQA				WebQuestions			
		Total	Question Overlap	Answer Overlap Only	No Overlap	Total	Question Overlap	Answer Overlap Only	No Overlap	Total	Question Overlap	Answer Overlap Only	No Overlap
Open book	RAG	44.5	70.7	34.9	24.8	56.8	82.7	54.7	29.2	45.5	81.0	45.8	21.1
	DPR	41.3	69.4	34.6	19.3	57.9	80.4	59.6	31.6	42.4	74.1	39.8	22.2
	FID	51.4	71.3	48.3	34.5	67.6	87.5	66.9	42.8	-	-	-	-
Closed book	T5-11B+SSM	36.6	77.2	22.2	9.4	-	-	-	-	44.7	82.1	44.5	22.0
	BART	26.5	67.6	10.2	0.8	26.7	67.3	16.3	0.8	27.4	71.5	20.7	1.6

Lewis et al. Overlap in Open-Domain Question Answering Datasets

Generalisation

Model	Natural Questions				TriviaQA			
	Total	Overlap	Comp -gen	Novel -entity	Total	Overlap	Comp -gen	Novel -entity

Unseen entity-relation pair

New entity

Liu et al. Challenges in Generalisation in Open-Domain Question Answering

Generalisation

Model		Natural Questions				TriviaQA			
		Total	Overlap	Comp -gen	Novel -entity	Total	Overlap	Comp -gen	Novel -entity
Parametric	T5-11B+SSM	36.59	**81.48**	17.47	12.56	-	-	-	-
	BART	26.54	76.34	5.88	3.35	26.78	78.38	11.37	10.09

Unseen entity-relation pair

New entity

Liu et al. Challenges in Generalisation in Open-Domain Question Answering

Generalisation

Model		Natural Questions				TriviaQA			
		Total	Overlap	Comp -gen	Novel -entity	Total	Overlap	Comp -gen	Novel -entity
Non-parametric	RAG	44.49	75.75	30.41	37.69	56.83	87.12	47.58	47.81
	FiD	**53.13**	78.85	**40.00**	**47.74**	**67.69**	**90.39**	**58.10**	**66.23**
	DPR	41.27	71.33	25.88	33.84	57.91	82.31	46.11	58.99
	RePAQ	47.26	78.61	34.21	36.85	52.06	89.08	42.95	38.38
Parametric	T5-11B+SSM	36.59	**81.48**	17.47	12.56	-	-	-	-
	BART	26.54	76.34	5.88	3.35	26.78	78.38	11.37	10.09

Liu et al. Challenges in Generalisation in Open-Domain Question Answering

Updateability

query (string)	answer (list)	date (string)
"Valentino Rossi plays for _X_."	[{ "wikidata_id": "Q1085474", "name": ["Monster Energy…	"2010"
"Valentino Rossi plays for _X_."	[{ "wikidata_id": "Q271812", "name": ["Ducati", "Ducati…	"2011"
"Valentino Rossi plays for _X_."	[{ "wikidata_id": "Q271812", "name": ["Ducati", "Ducati…	"2012"
"Valentino Rossi plays for _X_."	[{ "wikidata_id": "Q1085474", "name": ["Monster Energy…	"2013"
"Valentino Rossi plays for _X_."	[{ "wikidata_id": "Q1085474", "name": ["Monster Energy…	"2014"

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

Updateability

		2017 Test Set Acc.		2020 Test Set Acc.	
Train Set	Test-time Index	Closed-book	ATLAS	Closed-book	ATLAS
2017 answers	2017				
	2020				
2020 answers	2017				
	2020				

T5-XXL

Retrieval-augmented

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

Updateability

		2017 Test Set Acc.		2020 Test Set Acc.	
Train Set	Test-time Index	Closed-book	ATLAS	Closed-book	ATLAS
2017 answers	2017	12.1	Retrieval-augmented	2.9	
	2020	12.1		2.9	
2020 answers	2017	4.8		3.6	
	2020	4.8		3.6	

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

Updateability

		2017 Test Set Acc.		2020 Test Set Acc.	
Train Set	Test-time Index	Closed-book	ATLAS	Closed-book	ATLAS
2017 answers	2017	12.1	57.7	2.9	1.5
	2020	12.1	10.2	2.9	53.1
2020 answers	2017	4.8	50.1	3.6	4.2
	2020	4.8	3.5	3.6	60.5

Izacard et al. Few-shot Learning with Retrieval-Augmented Language Models

Retrieval-Augmented Models

- Control and updateability ✅

Retrieval-Augmented Models

- Control and updateability ✅
- Some degree of interpretability and provenance ✅

Retrieval-Augmented Models

- Control and updateability ✅
- Some degree of interpretability and provenance ✅
- More factual generations; fewer hallucinations ✅

Retrieval-Augmented Models

- Control and updateability ✅
- Some degree of interpretability and provenance ✅
- More factual generations; fewer hallucinations ✅
- Towards decoupling factual knowledge from parameters ✅

Retrieval-Augmented Models

- Control and updateability ✅
- Some degree of interpretability and provenance ✅
- More factual generations; fewer hallucinations ✅
- Towards decoupling factual knowledge from parameters ✅
- Need to generate and index **billions** of document representations ❌

Retrieval-Augmented Models

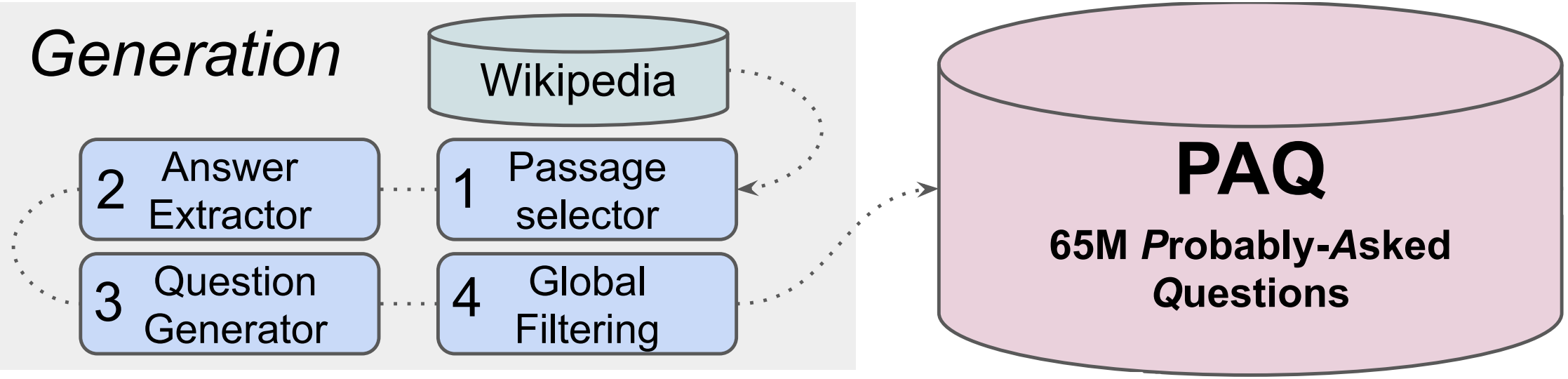
- Control and updateability ✅
- Some degree of interpretability and provenance ✅
- More factual generations; fewer hallucinations ✅
- Towards decoupling factual knowledge from parameters ✅
- Need to generate and index **billions** of document representations ❌
- The reader needs to read **hundreds** of documents ❌

Retrieval-Augmented Models

- Control and updateability ✅
- Some degree of interpretability and provenance ✅
- More factual generations; fewer hallucinations ✅
- Towards decoupling factual knowledge from parameters ✅
- Need to generate and index **billions** of document representations ❌
- The reader needs to read **hundreds** of documents ❌

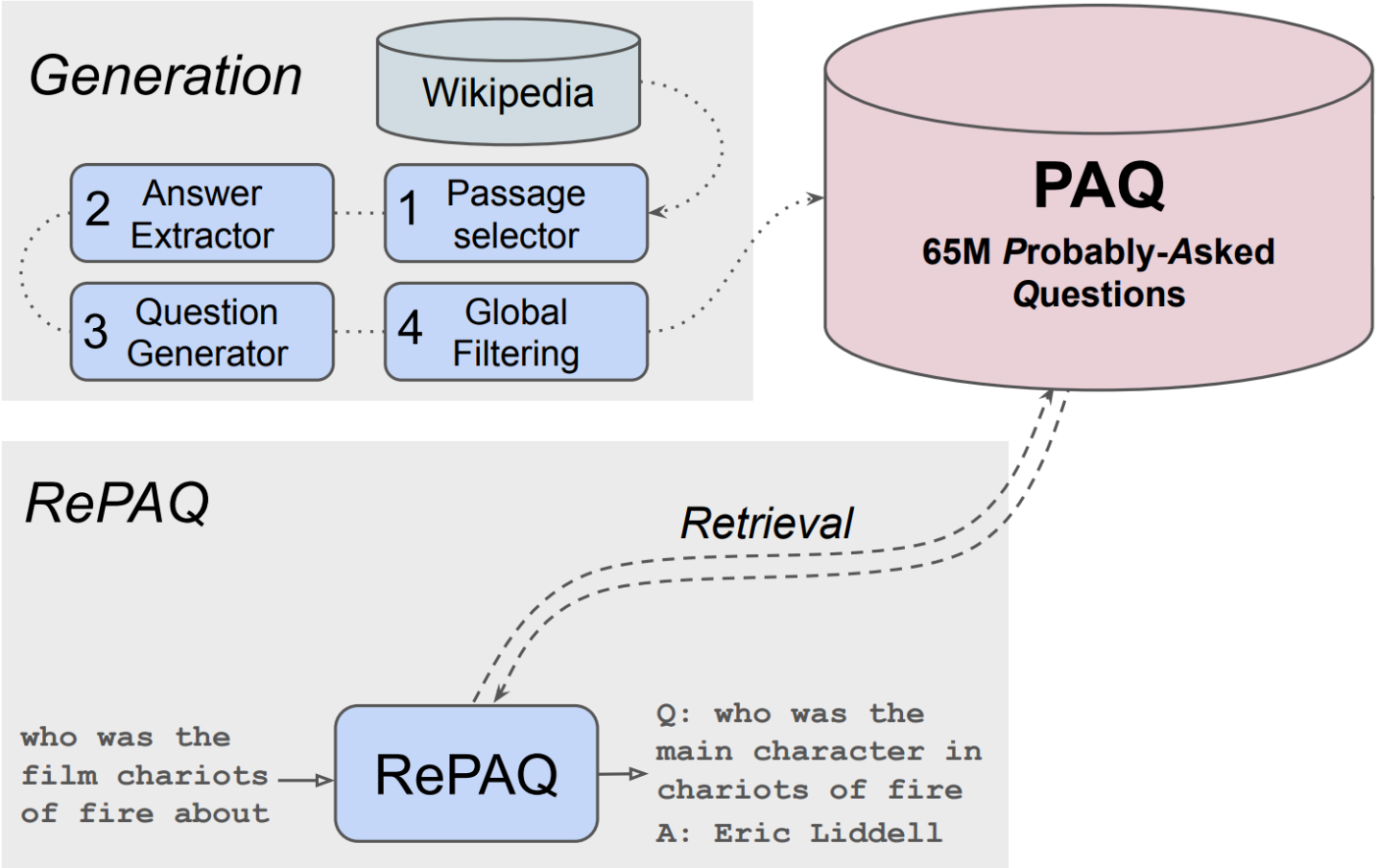
PAQ: A Condensed Wikipedia

We generated 65 million synthetic Question-Answer pairs ..



Lewis et al. PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them

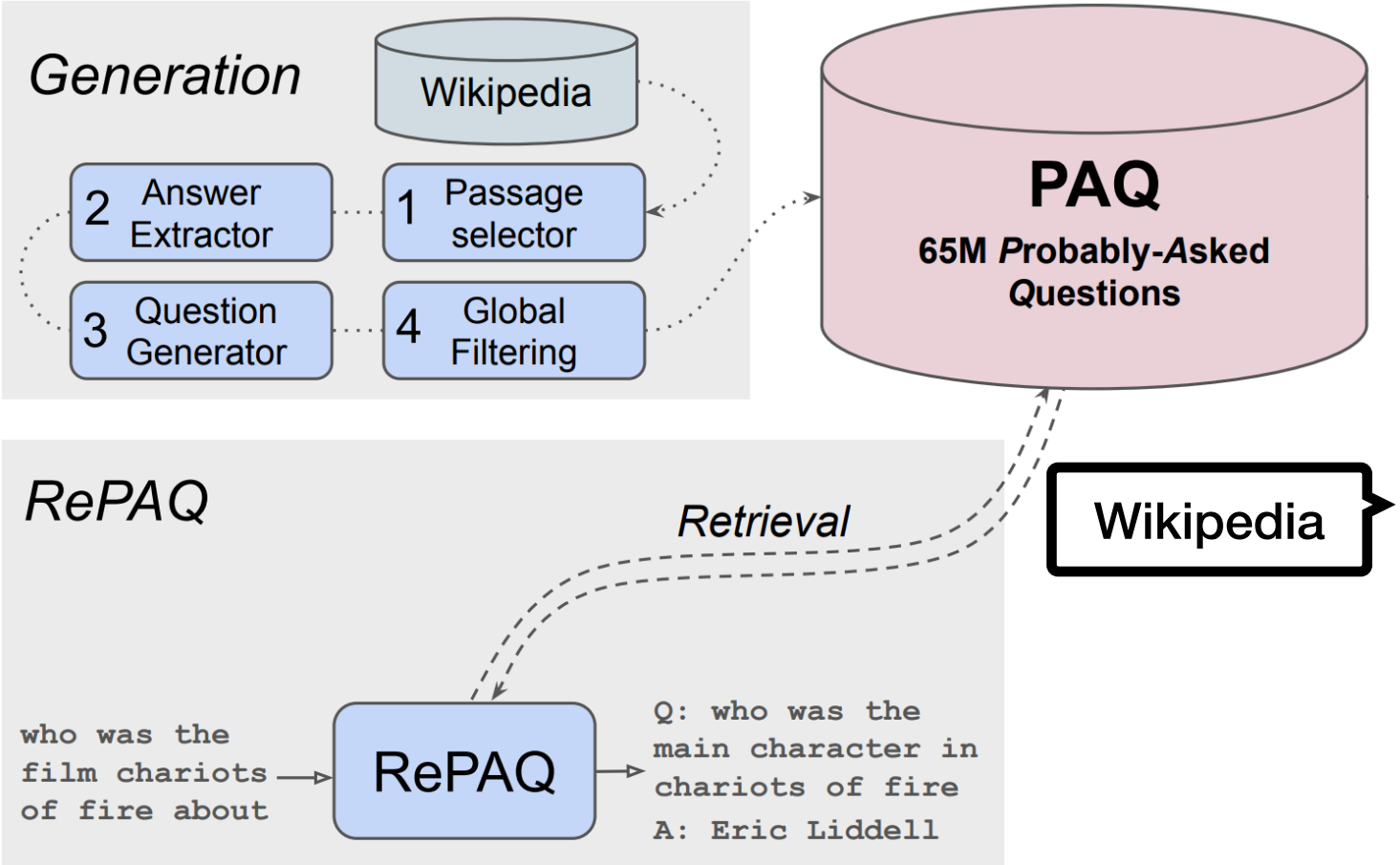
PAQ: A Condensed Wikipedia



and designed RePAQ, a retrieval-augmented QA model that uses the generated question-answer pairs as knowledge source.

Lewis et al. PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them

PAQ: A Condensed Wikipedia



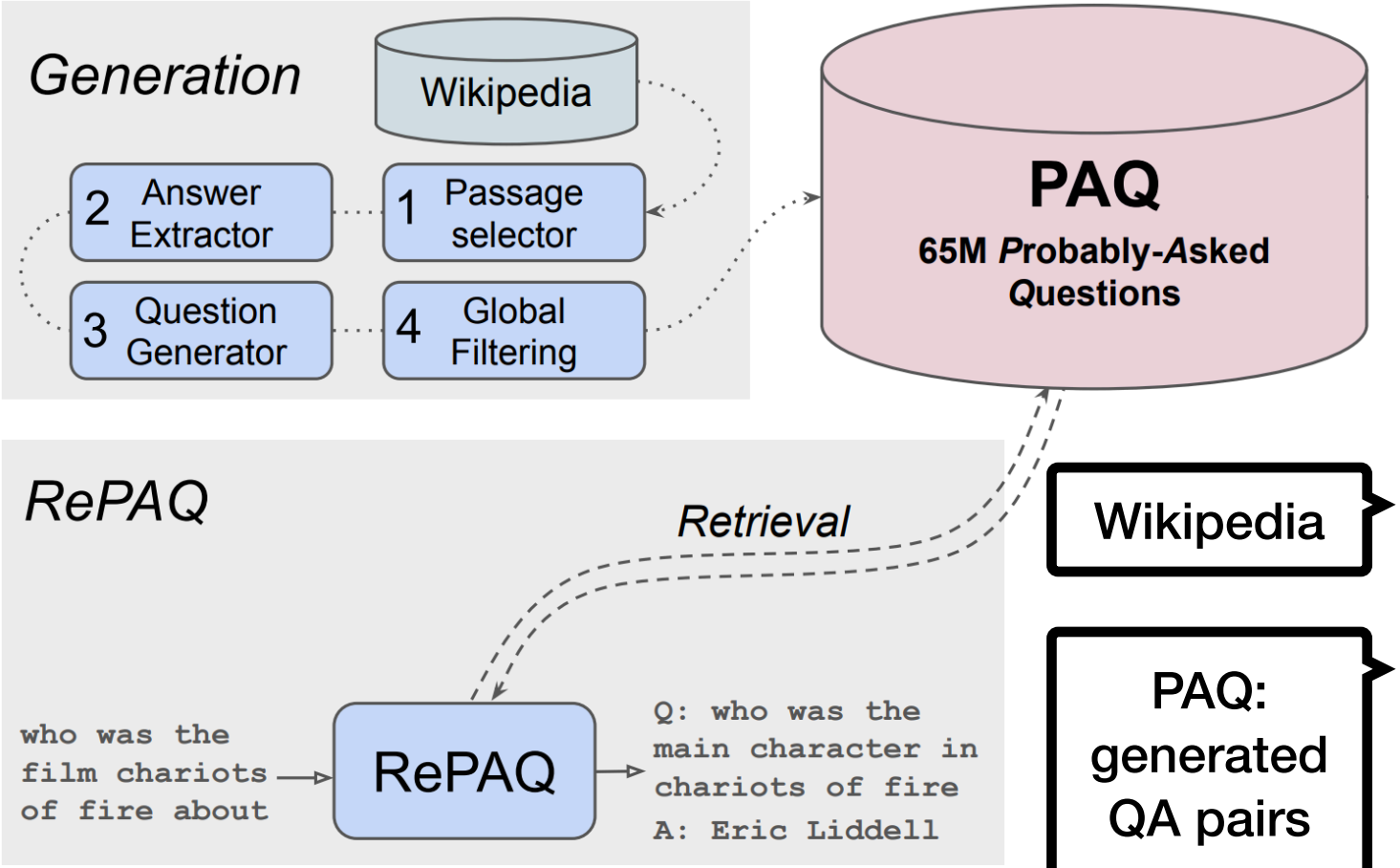
and designed RePAQ, a retrieval-augmented QA model that uses the generated question-answer pairs as knowledge source.

Wikipedia

Model	Retriever	Reranker	Exact Match	Q/sec
FiD-large	-	-	51.4	0.5
FiD-base	-	-	48.2	2

Lewis et al. PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them

PAQ: A Condensed Wikipedia



and designed RePAQ, a retrieval-augmented QA model that uses the generated question-answer pairs as knowledge source.

	Model	Retriever	Reranker	Exact Match	Q/sec
Wikipedia	FiD-large	-	-	51.4	0.5
	FiD-base	-	-	48.2	2
PAQ: generated QA pairs	RePAQ	base	-	40.9	1400
	RePAQ	large	-	41.2	1100
	RePAQ	xlarge	-	41.5	800
	RePAQ	base	base	45.7	55
	RePAQ	large	xlarge	46.2	10
	RePAQ	xlarge	xxlarge	47.6	6

Lewis et al. PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them

Efficient QA Challenge at NeurIPS 2020

Systems Under 500Mb

Rank	Model	Participant	Affiliation	Attempt Date	Accuracy	Size
1	uclnlp-fair-efficientqa-lg	uclnlp-efficientqa	University College London and Facebook AI Research	14/11/2020	33.44	336.23 MB

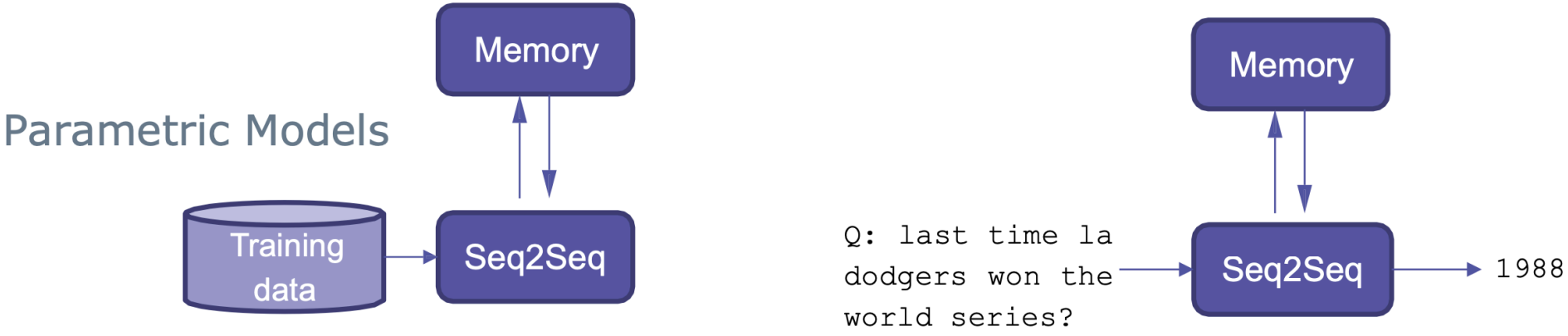
Smallest Systems Achieving 25% Accuracy

Rank	Model	Participant	Affiliation	Attempt Date	Accuracy	Size
1	uclnlp-fair-efficientqa	uclnlp-efficientqa	University College London and Facebook AI Research	07/11/2020	26.78	29.43 MB

Min et al. NeurIPS 2020 EfficientQA Competition: Systems, Analyses, and Lessons Learned

Efficient Memory-Augmented Transformers

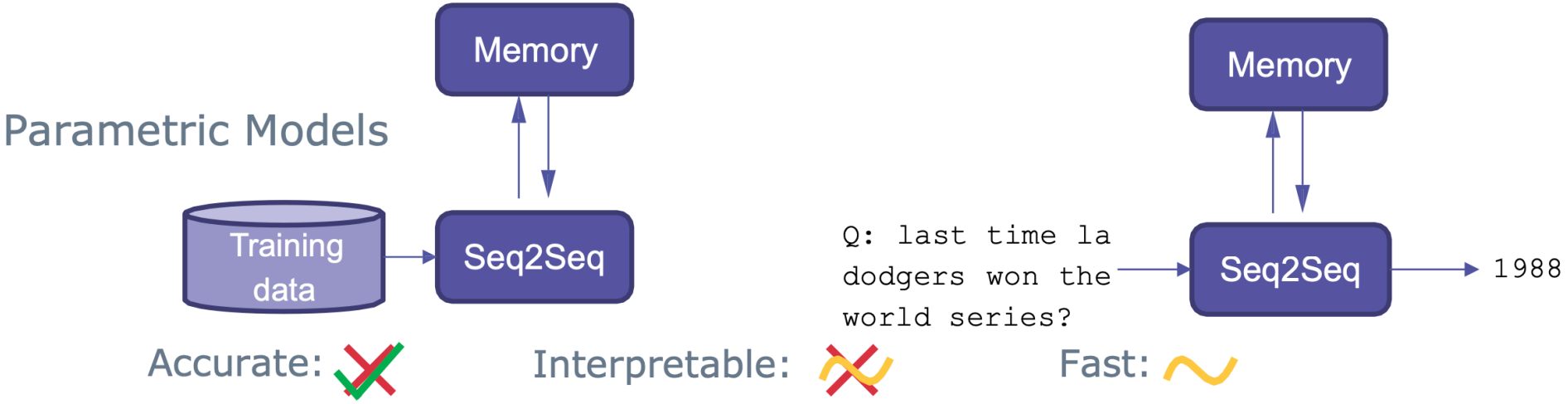
Generated QA pairs (for example) can be integrated in Transformer-based models:



Wu et al. An Efficient Memory-Augmented Transformer for Knowledge-Intensive NLP Tasks

Efficient Memory-Augmented Transformers

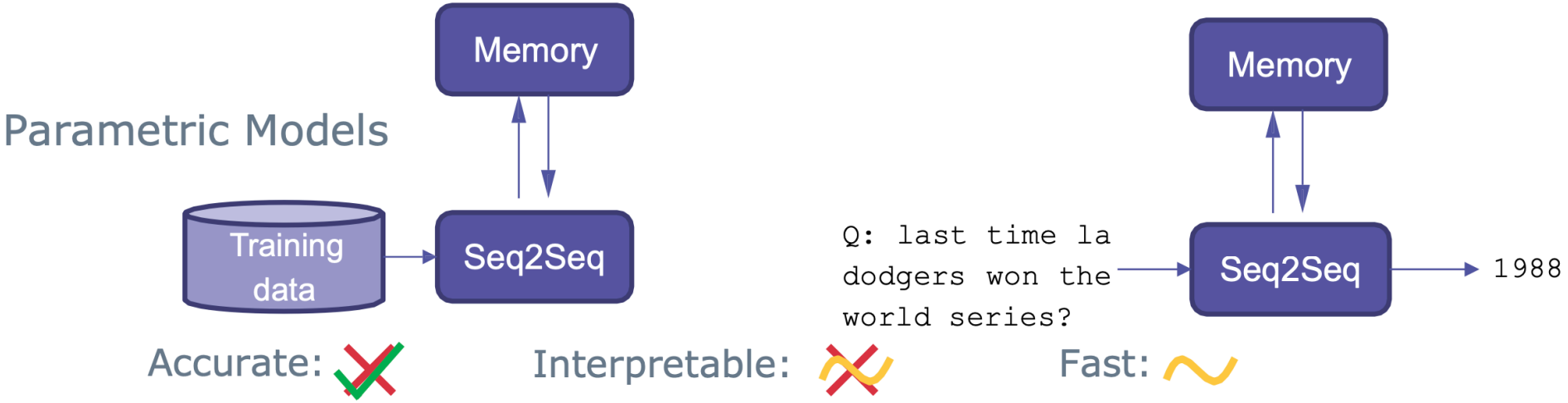
Generated QA pairs (for example) can be integrated in Transformer-based models:



Wu et al. An Efficient Memory-Augmented Transformer for Knowledge-Intensive NLP Tasks

Efficient Memory-Augmented Transformers

Generated QA pairs (for example) can be integrated in Transformer-based models:

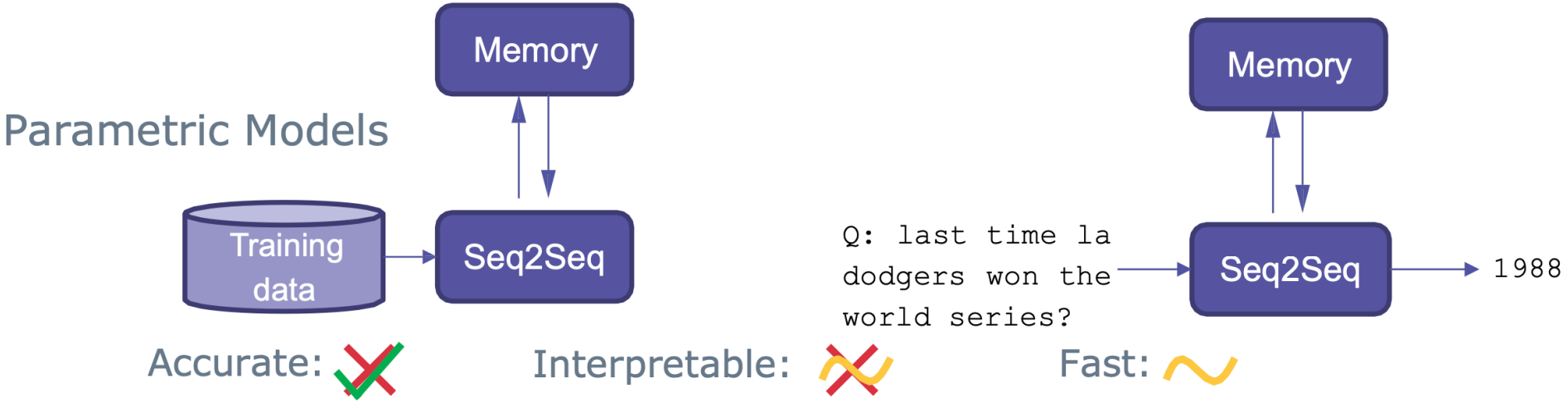


EMATs combine the strengths of Parametric and Retrieval-Augmented models:

Wu et al. An Efficient Memory-Augmented Transformer for Knowledge-Intensive NLP Tasks

Efficient Memory-Augmented Transformers

Generated QA pairs (for example) can be integrated in Transformer-based models:



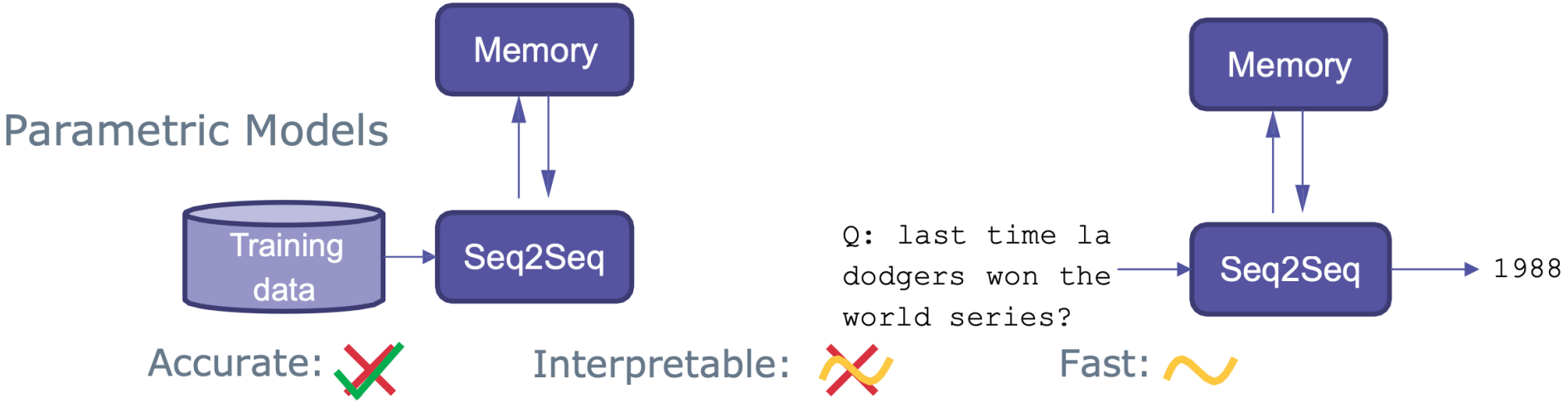
EMATs combine the strengths of Parametric and Retrieval-Augmented models:

- Knowledge encoded as a **differentiable key-value memory**

Wu et al. An Efficient Memory-Augmented Transformer for Knowledge-Intensive NLP Tasks

Efficient Memory-Augmented Transformers

Generated QA pairs (for example) can be integrated in Transformer-based models:

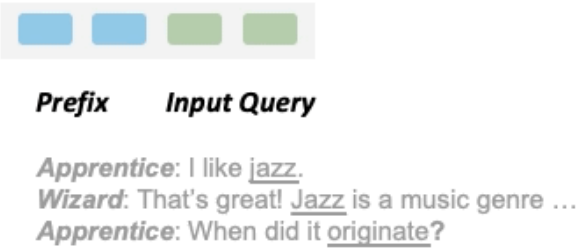


EMATs combine the strengths of Parametric and Retrieval-Augmented models:

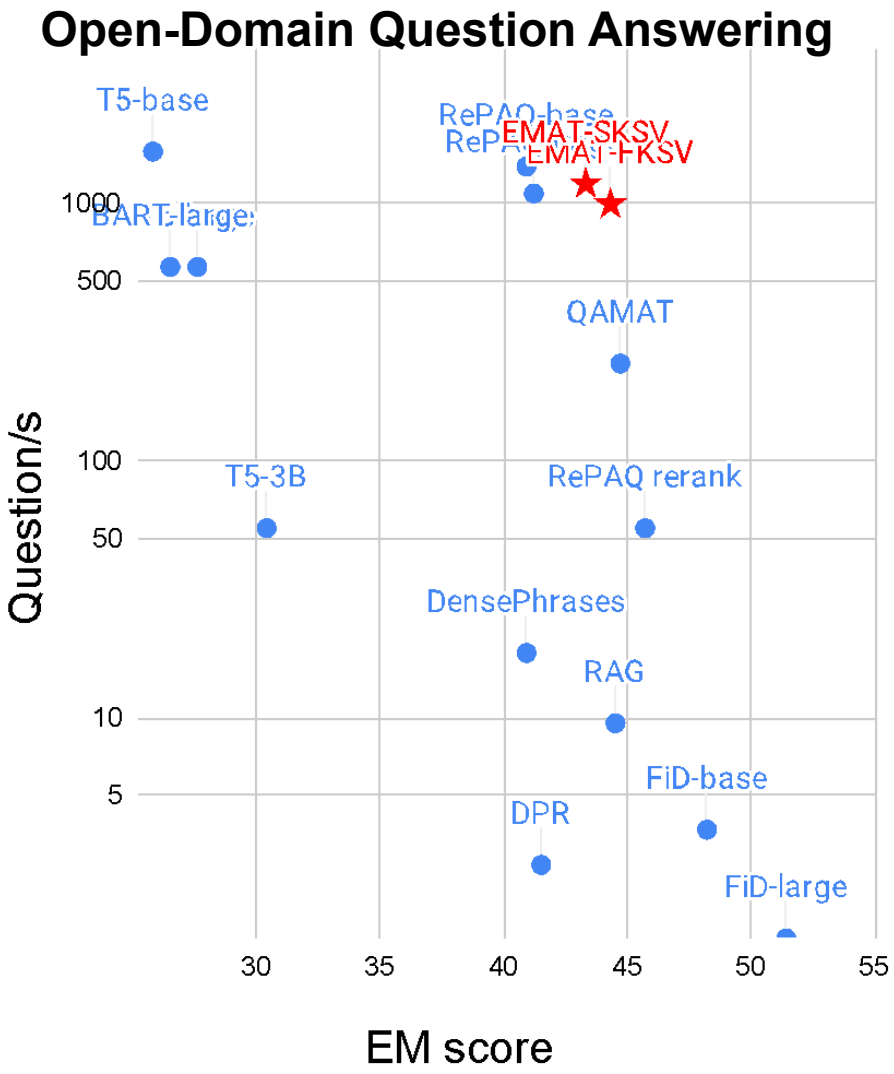
- Knowledge encoded as a **differentiable key-value memory**
- Memory retrieval via **fast MIPS search**

Wu et al. An Efficient Memory-Augmented Transformer for Knowledge-Intensive NLP Tasks

Efficient Memory-Augmented Transformers

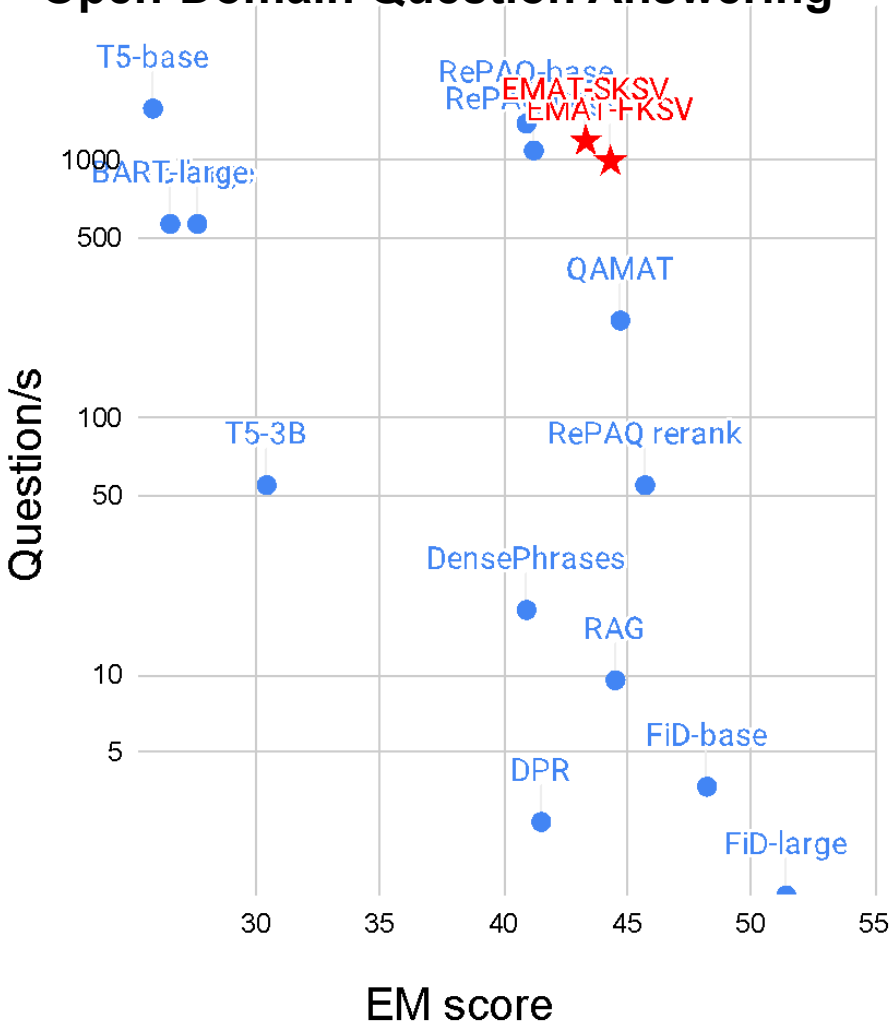


EMAT in Knowledge-Intensive Tasks

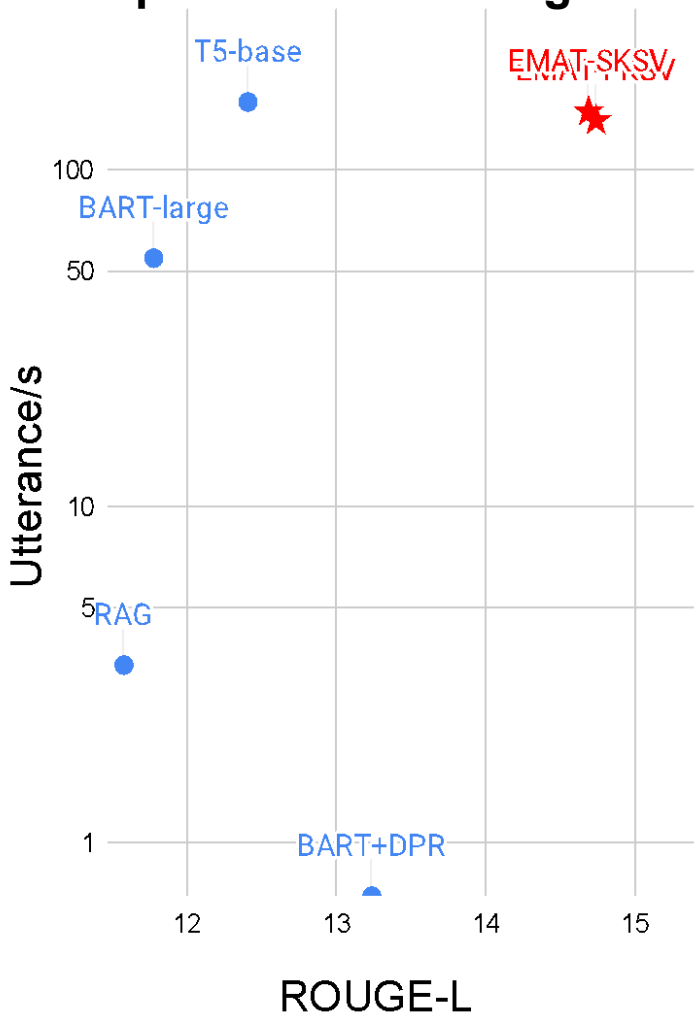


EMAT in Knowledge-Intensive Tasks

Open-Domain Question Answering

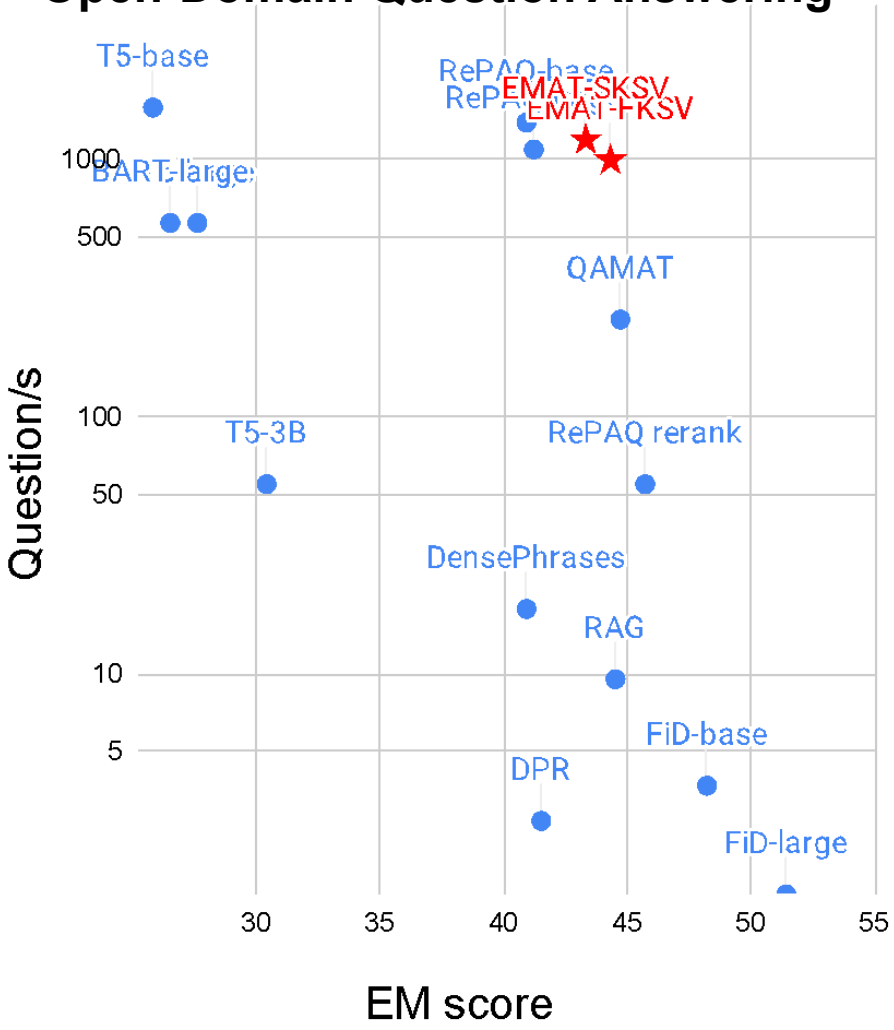


Open-Domain Dialogue

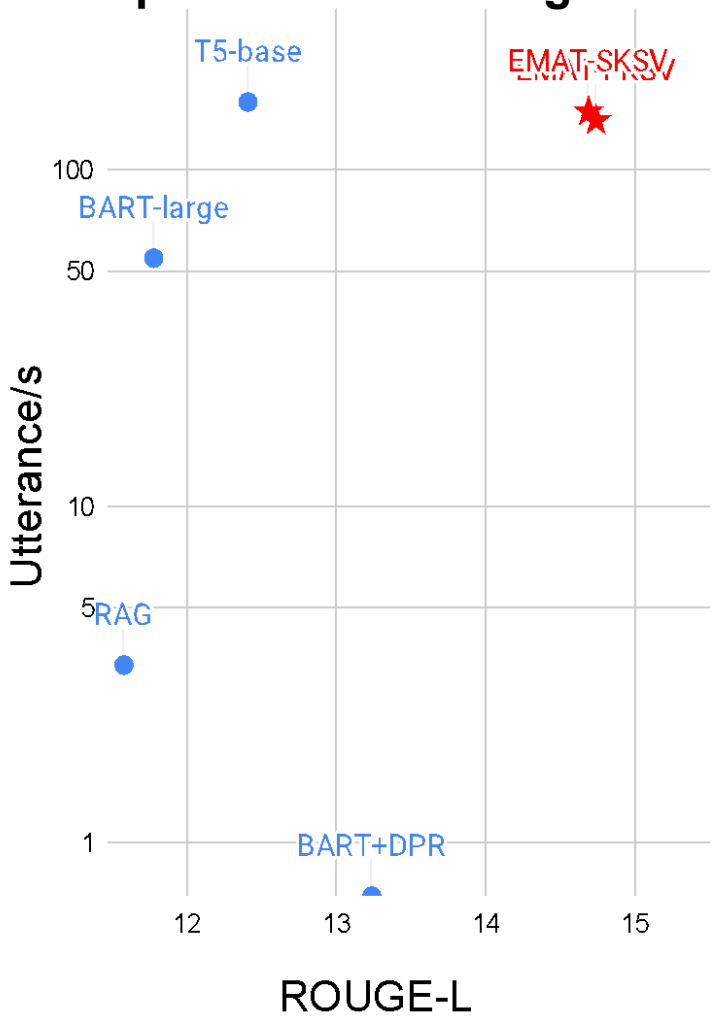


EMAT in Knowledge-Intensive Tasks

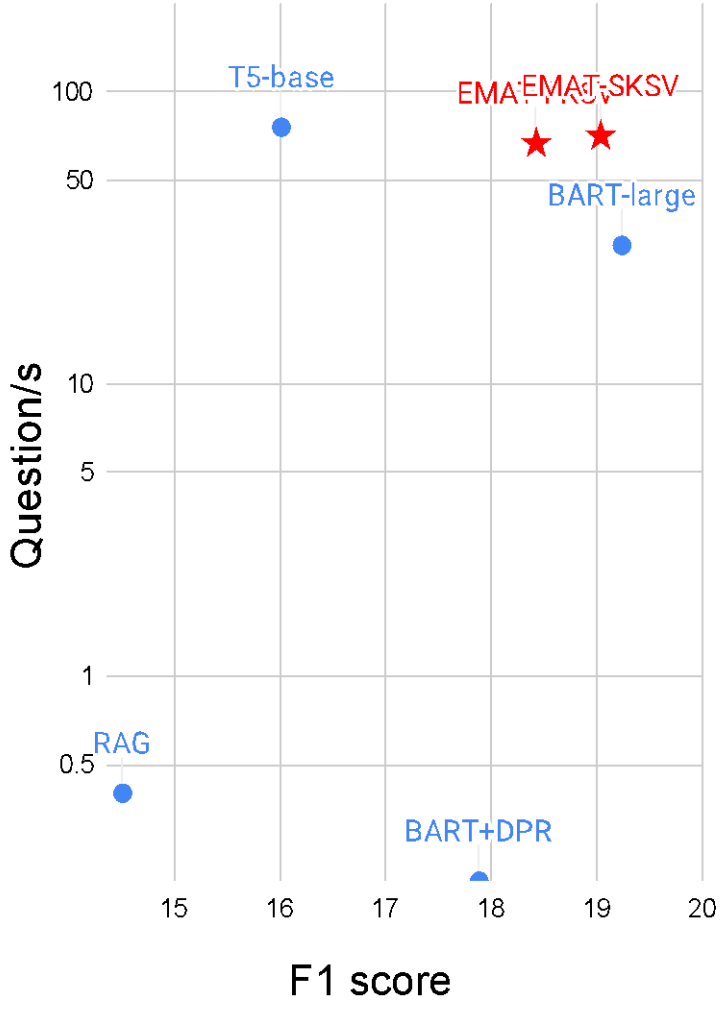
Open-Domain Question Answering



Open-Domain Dialogue



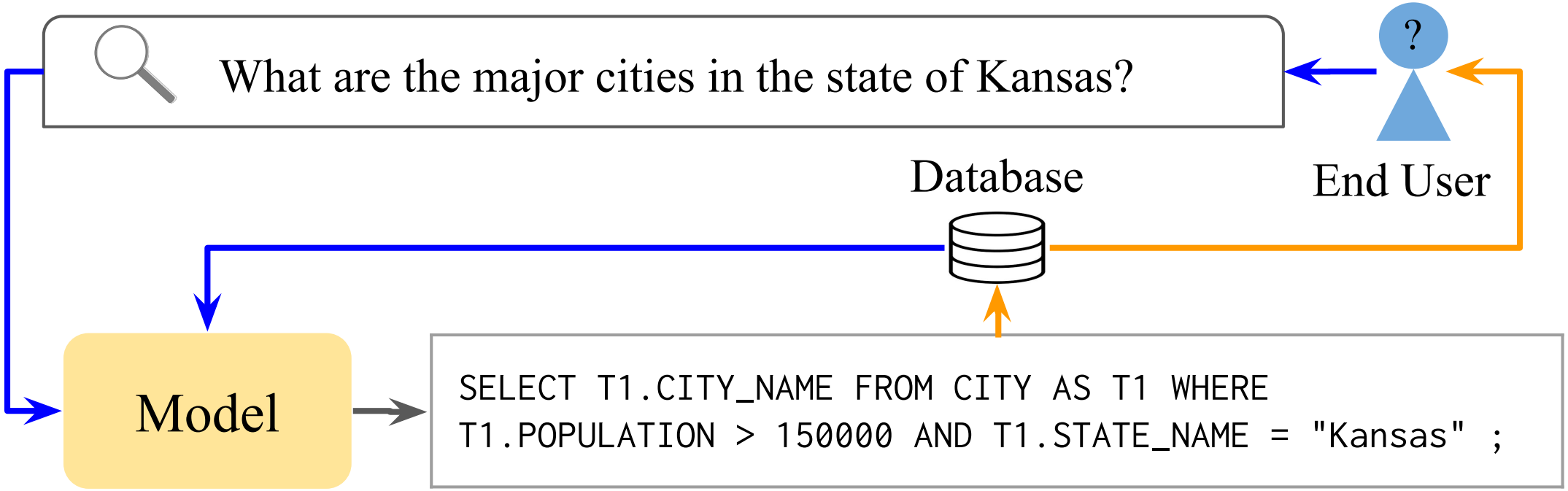
Long-Form Question Answering



Text-to-SQL/SPARQL

Text-to-SQL Systems

The framework for text-to-SQL systems: given a database schema and user utterance, the system outputs a corresponding SQL query to query the database system or the results.



Spider 1.0

Yale Semantic Parsing and Text-to-SQL Challenge

What is Spider?

Spider is a large-scale *complex and cross-domain* semantic parsing and text-to-SQL dataset annotated by 11 Yale students. The goal of the Spider challenge is to develop natural language interfaces to cross-domain databases. It consists of 10,181 questions and 5,693 unique complex SQL queries on 200 databases with multiple tables covering 138 different domains. In Spider 1.0, different complex SQL queries and databases appear in train and test sets. To do well on it, systems must *generalize well to not only new SQL queries but also new database schemas*.

Why we call it "Spider"? It is because our dataset is complex and cross-domain like a spider crawling across mutiple complex(with many foreign keys) nests(databases).

XLang Lab for building LM agents!

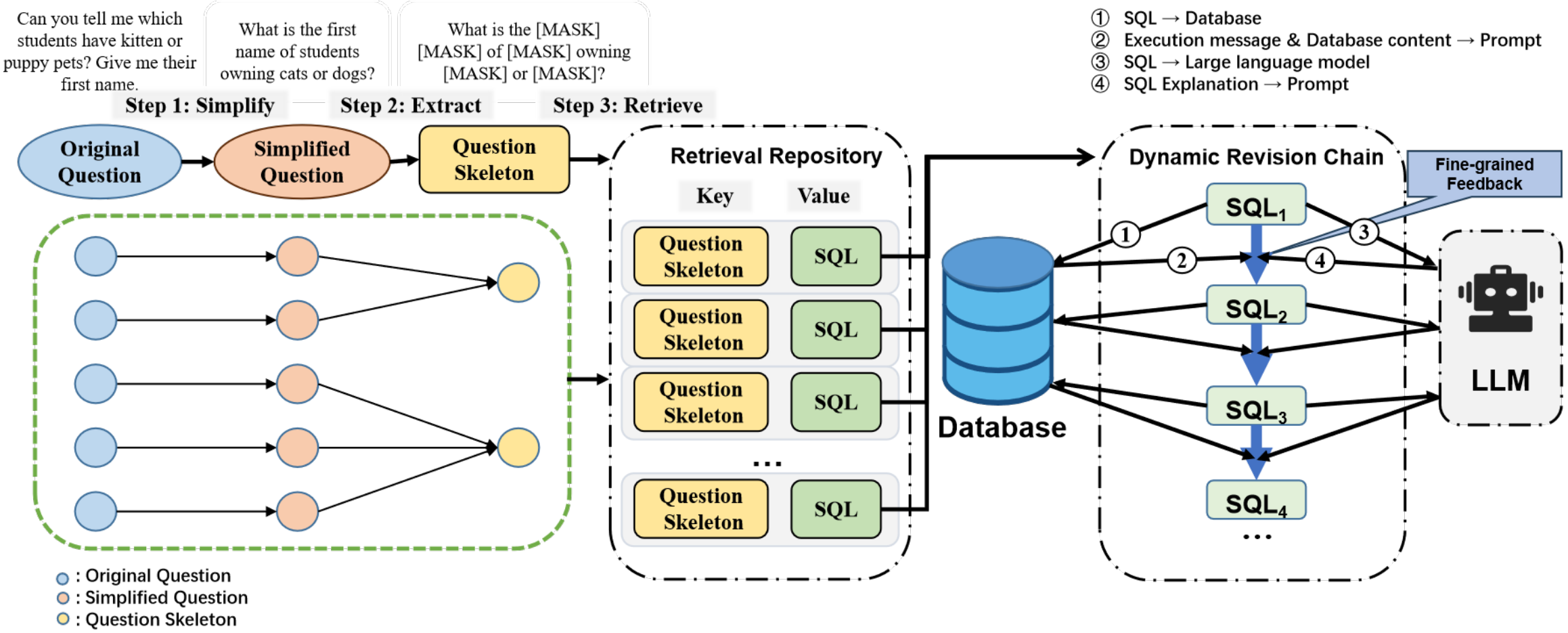
Spider Paper (EMNLP'18)

Leaderboard - Execution with Values

Our current models do not predict any value in SQL conditions so that we do not provide execution accuracies. However, we encourage you to provide it in the future submissions. For value prediction, your model should be able to 1) copy from the question inputs, 2) retrieve from the database content (database content is available), or 3) generate numbers (e.g. 3 in "LIMIT 3"). *Notice:* Test results after May 02, 2020 are reported on the new release (collected some annotation errors).

Rank	Model	Test
1 Nov 2, 2023	MiniSeek *Anonymous* Code and paper coming soon	**91.2**
1 Aug 20, 2023	DAIL-SQL + GPT-4 + Self-Consistency *Alibaba Group* (Gao and Wang et al.,'2023) code	**86.6**
2 Aug 9, 2023	DAIL-SQL + GPT-4 *Alibaba Group* (Gao and Wang et al.,'2023) code	**86.2**

The left half shows retrieval repository construction in three steps. The top three sentences are three specific instances each. The Green dashed box presents the training set. The right half is a dynamic revision chain with SQL queries generated by LLM iterations as nodes (green boxes). The output of steps 2 and 4 are collectively referred to as fine-grained feedback.



Language as Programs

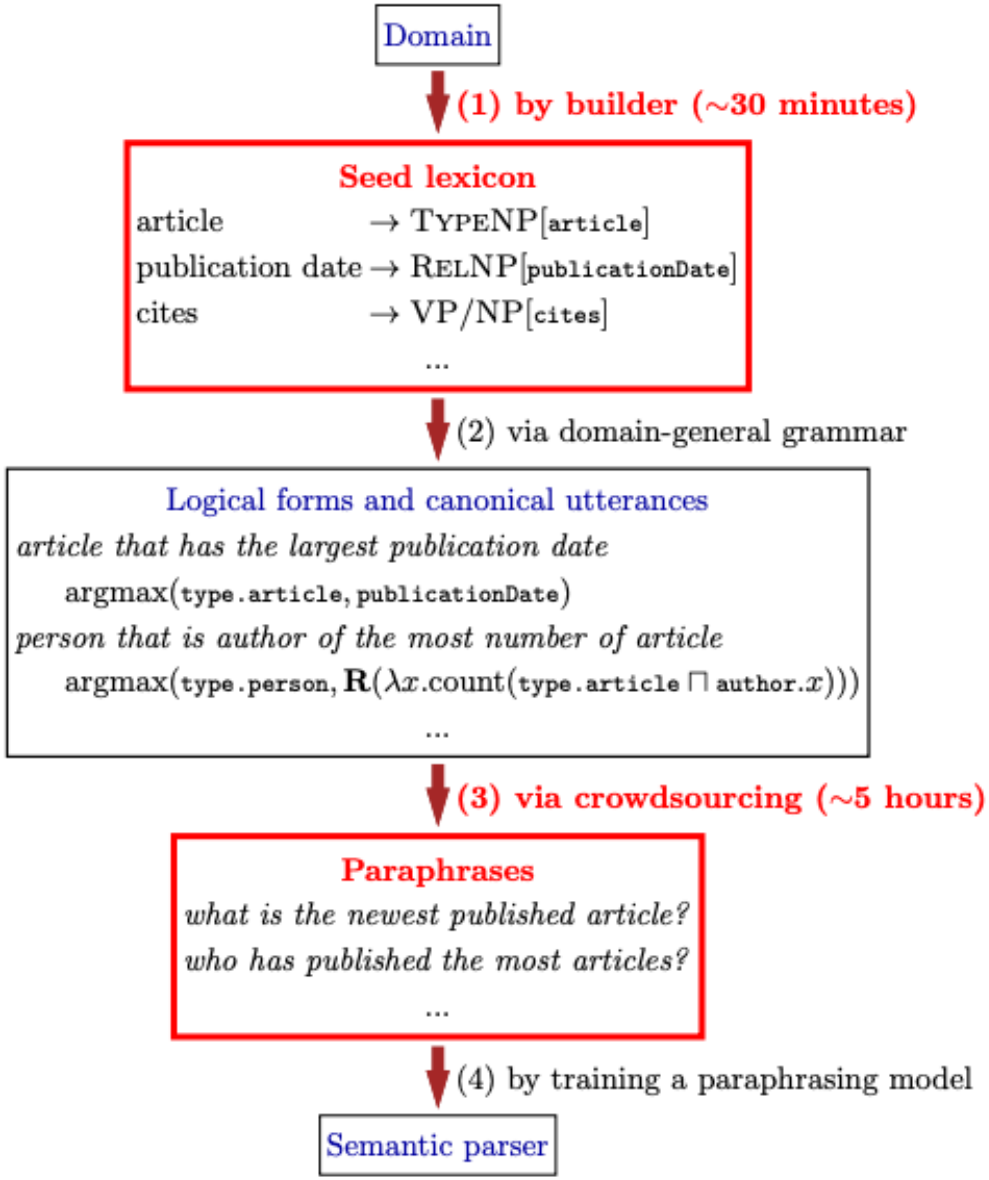
Remind me to buy milk after my last meeting on Monday.

↓ semantic parsing

$\texttt{Add}(\texttt{Buy}(\texttt{Milk}), \texttt{argmax}(\texttt{Meetings} \cap \texttt{HasDate}(\texttt{2016-07-18}), \texttt{EndTime}))$

↓ execute

[reminder added]



Building a Semantic Parser Overnight

Yushi Wang*
Stanford University
yushiw@cs.stanford.edu

Jonathan Berant*
Stanford University
joberant@stanford.edu

Percy Liang
Stanford University
pliang@cs.stanford.edu