



ifis

Institut für Informationssysteme
Technische Universität Braunschweig

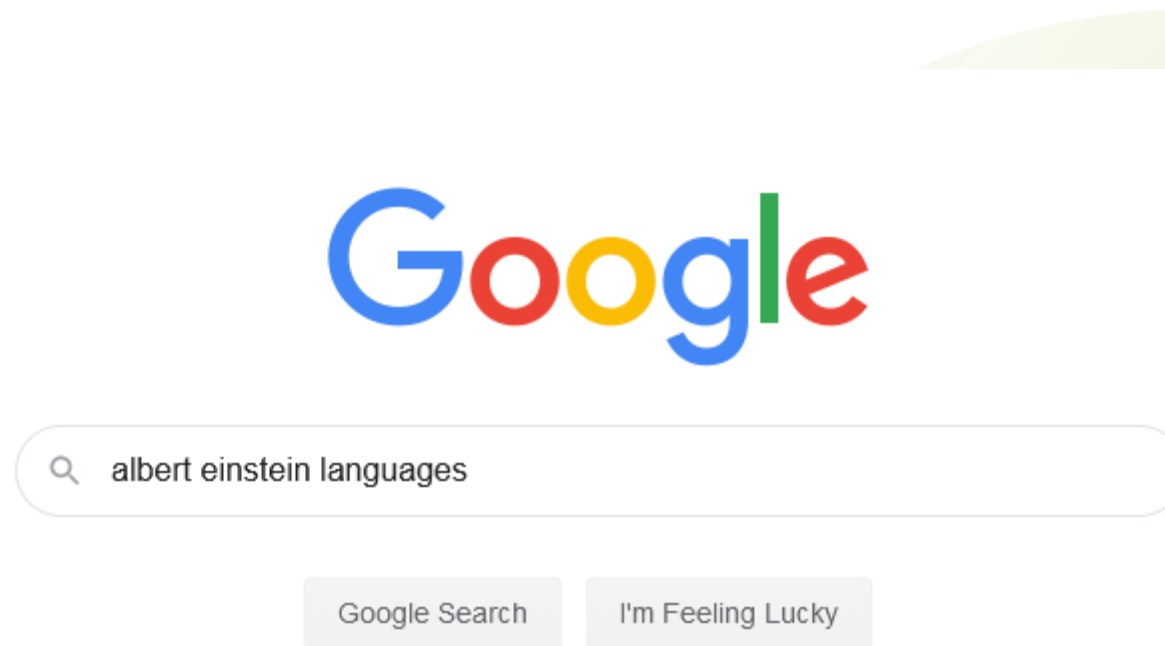
KnowlyBERT - Hybrid Query Answering over Language Models and Knowledge Graphs

Jan-Christoph Kalo, Leandra Fichtel, Philipp Ehler
and Wolf-Tilo Balke

Institut für Informationssysteme
Technische Universität Braunschweig



Entity-Centric Queries on Google





Entity-Centric Queries on Google

5 Answers



Gali Weinstein, Historian of modern physics

Answered Sep 22, 2018 · Upvoted by Thomas Wier, Assistant Professor of Linguistics at the Free University of Tbilisi. · Author has **394** answers and **3.9m** answer views

Until 1913 Einstein, a native German speaker, could not write, read and understand English. Afterwards, and when he lived in America, he wrote letters and the like either in impeccable English or in less idiomatic English. Leopold Infeld, Einstein's assistant, once said jokingly that Einstein's English contained about three hundred words pronounced in a peculiar way. Banesh Hoffmann later recalled that when Einstein found himself up against a seemingly insuperable difficulty, he would stand up, put his pipe on the table, and say in his quaint English, "I will a little tink"!

Einstein wrote his matriculation examination essay in French, entitled "My Future Plans", written at the age of seventeen but he later said, "my French has never been very good".

[www.quora.com › How many languages did Ein... › Diese Seite übersetzen](#)

[How many languages did Einstein speak? - Quora](#)

5 Antworten

12.03.2016 - Until 1913 Einstein, a native German speaker, could not write, read and understand ... Barbara Wolff, former Archivist at the **Albert Einstein Archives** (1995-2016).

[www.openculture.com › 2013/03 › listen_as_alb... › Diese Seite übersetzen](#)

[Albert Einstein Reads 'The Common Language of Science ...](#)

21.03.2013 - "Is there no thinking without the use of language," asks Einstein, "namely in concepts and concept-combinations for which words need not ...

[dipc.ehu.es › digitalak › orriak › english › people › Diese Seite übersetzen](#)

[Albert Einstein - DIPC](#)

In accepting the invitation, Einstein had explained his problems with languages to Rey Pastor: "I will accept your invitation on condition that I limit my lectures to ...

[www.huffpost.com › entry › abert-einstein-facts,... › Diese Seite übersetzen](#)

[10 Surprising Facts About Albert Einstein | HuffPost](#)

17.10.2013 - We all know the basics about **Albert Einstein**. He was one of the greatest scientists of all time. He came up with the theory of relativity. He was ...



Querying Wikidata

- **SPARQL query** on Wikidata: `SELECT ?languages`

`WHERE`

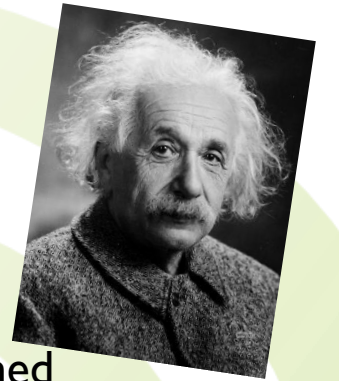
`{`

`wd:Q937 wdt:P1412 ?languages.`

`}`

Albert Einstein

languages spoken, written or signed



Answers

German

English

- **French** is not a language which can be found in Wikidata



Recent Advances in NLP

- **Language Models** contain large amounts of **knowledge**

Language Models as Knowledge Bases?

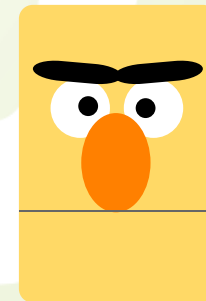
Fabio Petroni¹ Tim Rocktäschel^{1,2} Patrick Lewis^{1,2} Anton Bakhtin¹
Yuxiang Wu^{1,2} Alexander H. Miller¹ Sebastian Riedel^{1,2}

¹Facebook AI Research

²University College London

{fabio.petroni, rockt, plewis, yolo, yuxiangwu, ahm, sriedel}@fb.com

Hey Bert,
do you know which languages were
spoken by Albert Einstein?

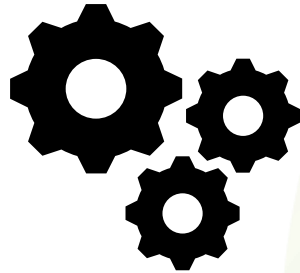
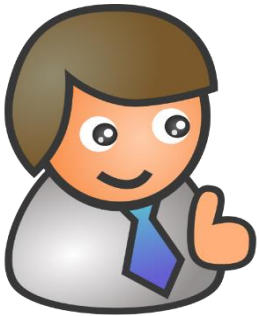




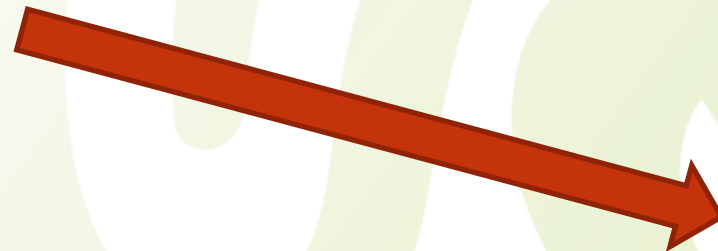
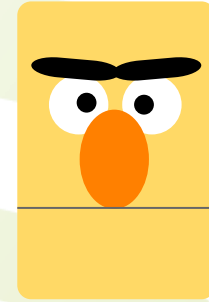
KnowlyBERT

- KnowlyBERT is a **hybrid query system** combining a language model and a knowledge graph

```
SELECT ?languages
WHERE
{
  wd:Q937 wdt:P1412 ?languages.
}
```



hybrid system

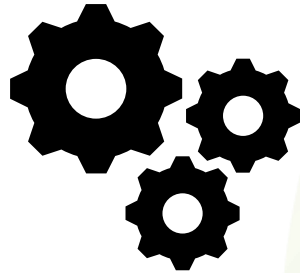
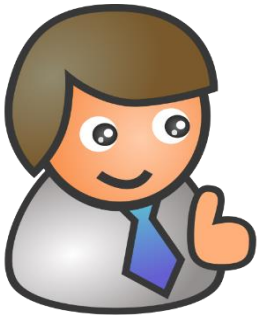




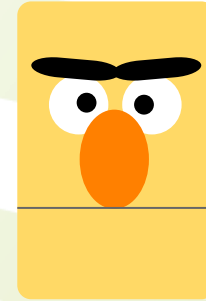
KnowlyBERT

- Hybrid system passes SPARQL directly query to Wikidata

```
SELECT ?languages
WHERE
{
  wd:Q937 wdt:P1412 ?languages.
}
```



hybrid system



```
SELECT ?languages
WHERE
{
  wd:Q937 wdt:P1412 ?languages.
}
```

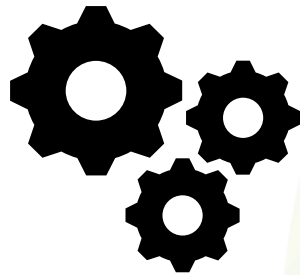
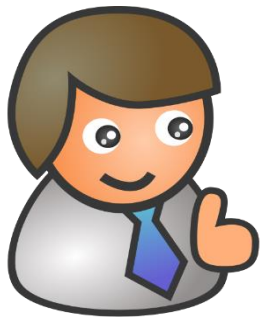




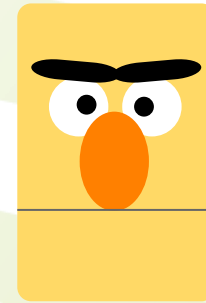
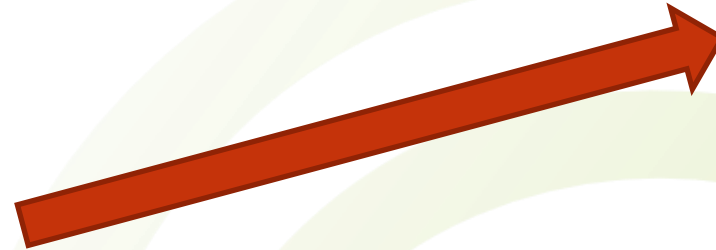
KnowlyBERT

- Wikidata supplies correct but **incomplete** answers

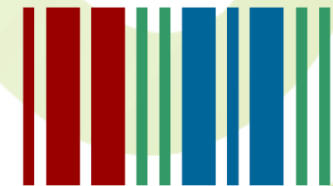
```
SELECT ?languages
WHERE
{
  wd:Q937 wdt:P1412 ?languages.
}
```



hybrid system



Answer
German
English



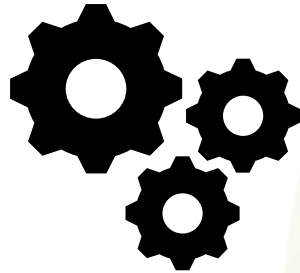
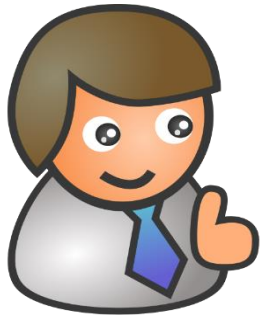
WIKIDATA



KnowlyBERT

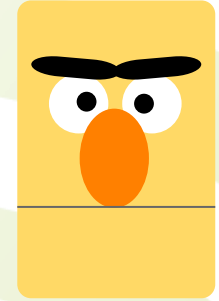
- Translate SPARQL query to sentence

```
SELECT ?languages
WHERE
{
  wd:Q937 wdt:P1412 ?languages.
}
```



hybrid system

„Albert Einstein used to communicate in [MASK].“



Answer
German
English

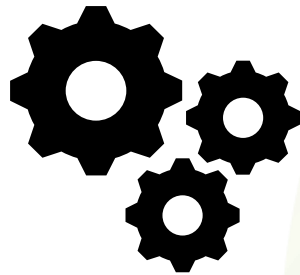
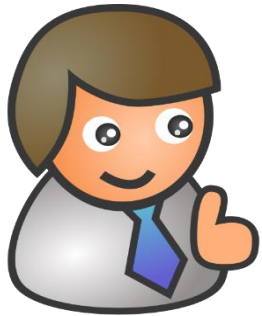




KnowlyBERT

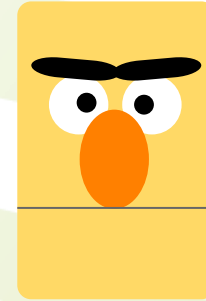
- Many incorrect answers
- But also **all correct answers**

```
SELECT ?languages
WHERE
{
  wd:Q937 wdt:P1412 ?languages.
}
```



hybrid system

Answer	Conf.
languages	-0.86
English	- 2,06
the	-2.11
French	- 2,16
German	- 2,61
Russian language	- 3,18
Japanese	- 3,30
Latin	- 3,66
it	- 3,70



Answer
German
English





Multiple Filtering Technologies

- Multiple Translations for Queries
 - SPARQL queries are translated into multiple sentences
- Entity Linking
 - Eliminate answers which cannot be mapped to Wikidata entities
- Type filter for entities
 - Domain and ranges for predicates are used to filter possible answers
- Popularity filter for entities
 - Eliminate very rare entities from answers
- Thresholding techniques
 - Delete low ranked answers

Answer	Conf.
languages	- 0.86
English	- 2.06
the	- 2.11
French	- 2.16
German	- 2.61
Russian language	- 3.18
Japanese	- 3.30
Latin	- 3.66
it	- 3.70



Multiple Filtering Technologies

- Multiple Translations for Queries
 - SPARQL queries are translated into multiple sentences
- Entity Linking
 - Eliminate answers which cannot be mapped to Wikidata entities
- Type filter for entities
 - Domain and ranges for predicates are used to filter possible answers
- Popularity filter for entities
 - Eliminate very rare entities from answers
- Thresholding techniques
 - Delete low ranked answers

Answer	Conf.
languages	- 0.86
English	- 2.06
the	- 2.11
French	- 2.16
German	- 2.61
Russian language	- 3.18
Japanese	- 3.30
Latin	- 3.66
it	- 3.70



Evaluation Setup

- **6649 randomly created queries** on Wikidata using 41 different properties
 - Subject queries: (?person, language spoken, English)
 - Object queries: (Albert Einstein, language spoken, ?language)
- Relation Extraction on Wikipedia Abstracts
 - Only queries whose answers can be found in the abstracts
- Knowledge Graph Embeddings using HoLE
 - Trained on all relevant Wikidata triples





Evaluation Results

Evaluation	Parameter	#Queries	Relation Extraction		Knowledge Graph Embedding		KnowlyBERT	
			Precision	Recall	Precision	Recall	Precision	Recall
Total		6649	17.5	17.6	<0.1	16.2	47.5	10.1
Cardinality	1-1	400	5,5	5,5	<0.1	20.2	16.9	3.0
	1-n	3756	18.8	17.4	<0.1	11.5	55.0	13.7
	n-m	2493	16.4	19.8	<0.1	22.6	36.0	5.9
Query Type	(s,p,?)	4029	37.5	17.3	<0.1	20.5	51.0	16.5
	(?,p,o)	2620	6.9	17.9	<0.1	9.5	10.5	0.3
Words	single	2474	39.6	13.9	<0.1	21.1	59.6	25.9
	multi	4175	13.0	19.7	<0.1	13.2	11.4	0.8
#Results	1	3797	40.5	13.2	<0.1	15.8	51.3	17.4
	2-10	1367	18.7	20.5	<0.1	20.4	37.0	4.9
	11-100	796	7.4	30.7	0.2	24.7	15.8	0.1
	>100	989	5.7	18.2	<0.1	4.8	<0.1	<0.1



Evaluation Results

Evaluation	Parameter	#Queries	Relation Extraction		KnowlyBERT	
			Precision	Recall	Precision	Recall
Total		6649	17.5	17.6	47.5	10.1
Query Type	(s,p,?)	4029	37.5	17.3	51.0	16.5
	(?,p,o)	2620	6.9	17.9	10.5	0.3



Results by Property

- **Locations** and **Languages** are predicted pretty good

Relation	Precision	Recall
Country	97.4	51.0
Birthplace	73.3	11.5
Instance of	<0.1	<0.1
Capital	15.4	3.0
Field of work	45.1	7.8
Native language	100.0	74.3
employer	100.0	0.6
Headquarter	56.8	13.2
Subclass of	16.7	<0.1
Instrument	<0.1	<0.1
Language spoken	45.8	17.7



Results by Property

- **Class Membership and Class Hierarchy** cannot be extracted at all

Relation	Precision	Recall
Country	97.4	51.0
Birthplace	73.3	11.5
Instance of	<0.1	<0.1
Capital	15.4	3.0
Field of work	45.1	7.8
Native language	100.0	74.3
employer	100.0	0.6
Headquarter	56.8	13.2
Subclass of	16.7	<0.1
Instrument	<0.1	<0.1
Language spoken	45.8	17.7



Conclusion

- KnowlyBERT shows **higher precision** than Relation Extraction and Knowledge Graph Completion techniques
- Combining **language models** and **knowledge graphs** offers high potential
 - Language models cover a plethora of knowledge
- It is unclear which knowledge is covered by language models
 - **Coverage** for some relations is very low



Future Work

- Which knowledge is **covered by language models**?
- How do we pose **complex SPARQL queries** to language models?
- Is there **other ways** how language models may support knowledge graphs



Take Away Message

Combining **implicit** and **explicit** knowledge representation offers great potential for future research



@janckalo



kalo@ifis.cs.tu-bs.de



<https://github.com/JanKalo/KnowlyBERT>