

# A Large-Scale Dataset for Known-Item Question Performance Prediction

---

QPP++ 2023, 6. April 2023, Dublin, Ireland



**Maik  
Fröbe<sup>1</sup>**



**Eric Oliver  
Schmidt<sup>2</sup>**



**Matthias  
Hagen<sup>1</sup>**



<sup>1</sup>  
**FRIEDRICH-SCHILLER-  
UNIVERSITÄT  
JENA**



<sup>2</sup>  
**MARTIN-LUTHER-UNIVERSITÄT  
HALLE-WITTENBERG**

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Motivation

- Query performance prediction experiments run often on TREC benchmarks
- Many currently used TREC benchmarks focus on informational queries, e.g.:
  - Robust04, MS MARCO, ...

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Motivation

- ❑ Query performance prediction experiments run often on TREC benchmarks
- ❑ Many currently used TREC benchmarks focus on informational queries, e.g.:
  - Robust04, MS MARCO, ...



# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Motivation

- ❑ Query performance prediction experiments run often on TREC benchmarks
- ❑ Many currently used TREC benchmarks focus on informational queries, e.g.:
  - Robust04, MS MARCO, ...



- ❑ Many queries submitted to search engines are navigational  
[Broder'02]
- ❑ Known-item and re-finding queries, can be challenging  
[Hagen'15, Arguello'21]
  - Searchers might be unable to recall suitable identifier for the item

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Motivation

Me, trying to recall the name of a movie...



It is on the tip of my tongue.

# A Large-Scale Dataset for Known-Item Question Performance Prediction

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.



# A Large-Scale Dataset for Known-Item Question Performance Prediction

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.



**TREC 2023 Tip-of-the-Tongue Track** <https://trec-tot.github.io/>  
[Arguello'21, Bhargav'22]

The phenomenon of failing to retrieve something from memory, combined with partial recall and the feeling that retrieval is imminent.

# A Large-Scale Dataset for Known-Item Question Performance Prediction

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.



**TREC 2023 Tip-of-the-Tongue Track** <https://trec-tot.github.io/>  
[Arguello'21, Bhargav'22]

The phenomenon of failing to retrieve something from memory, combined with partial recall and the feeling that retrieval is imminent.

**Tip-of-the-tongue Known-Item Retrieval Dataset (TOT-KIR):**  
[Bhargav'22]

- ❑ 15,000 questions with linked known items
- ❑ Focus: Movies and books
- ❑ Extracted from [reddit.com/r/tipofmytongue/](https://www.reddit.com/r/tipofmytongue/)
- ❑ Crawled: 2017—2020



# A Large-Scale Dataset for Known-Item Question Performance Prediction

Goal: Find the known-item

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.



 1   7

7 comments

Add a Comment

 SassyPants5 27 • 53 Seconds later




Poison Ivy

 7   Reply

 sonz82 18 • 5 Minutes later

Wow that was fast, yes this is the one! Thank you!



Solved!

 1   Reply

 WhatIsThisBot 2/∞ • 2 Days later

You have been given one point for this answer. Thanks for contributing!

sonz82 awarded to SassyPants5 (6)

 1   Reply

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

1 7

7 comments

Add a Comment

SassyPants5 27 • 53 Seconds later

Poison Ivy

7 Reply

sonz82 18 • 5 Minutes later

Wow that was fast, yes this is the one! Thank you!

Solved!

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

↑ 1 ↓ 7

7 comments

Add a Comment

SassyPants5 27 • 53 Seconds later

Poison Ivy

↑ 7 ↓ Reply

sonz82 18 • 5 Minutes later

Wow that was fast, yes this is the one! Thank you!

Solved!

[late-2000s] Video clip of Thom Yorke talking about Radiohead album

Around the time "In Rainbows" came out, I remember seeing a video clip of Thom Yorke talking to the viewer. He did something weird and offbeat immediately, said something weird, maybe a little dance or moving his head around or something. Thanks in advance!

↑ 2 ↓ 6

6 comments

Add a Comment

< [deleted] 27 • 3 Days later

Was it [this one](#) from "on the Big Fat Quiz of the Year"?

↑ 1 ↓ Reply

savorie 18 • 4 Days later

Solved! OMG!! THIS WAS IT!

how did you figure it out?

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

1 up 7 down

7 comments  
[Add a Comment](#)

SassyPants5 27 • 53 Seconds later

Poison Ivy

7 up 1 down Reply

sonz82 18 • 5 Minutes later

Wow that was fast, yes this is the one! Thank you!

Solved!

[late-2000s] Video clip of Thom Yorke talking about Radiohead album

Around the time "In Rainbows" came out, I remember seeing a video clip of Thom Yorke talking to the viewer. He did something weird and offbeat immediately, said something weird, maybe a little dance or moving his head around or something. Thanks in advance!

2 up 6 down

6 comments  
[Add a Comment](#)

[deleted] 27 • 3 Days later

Was it **this one** from "on the Big Fat Quiz of the Year"?

1 up 1 down Reply

savorie 18 • 4 Days later

Solved! OMG!! THIS WAS IT!

how did you figure it out?

[Video] Recreation of Greys Anatomy surgery scene

These people recreated a scene in Greys Anatomy where someone puts on the surgeons gloves for them, then they operate and the patient dies and they continue to try and revive them. The caption is something along the lines of "Greys Anatomy be like..."

1 up 1 down 1 comment

1 comments  
[Add a Comment](#)

PhDTotoro 13 • 7 Days later

I remember this one. But I don't think it was Grey's Anatomy related.

1 up 1 down Reply

not solved!

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

The image shows three examples of social media posts with their respective comments. Each post includes a title, a description, and interaction metrics (upvotes, downvotes, replies). The first post is about a 1990s movie, the second is about a Radiohead video clip, and the third is about a Greys Anatomy surgery scene. The comments show users providing answers or reactions, with some indicating when they solved the question.

[Movie] From the 1990s?  
I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.  
↑ 1 ↓ 7

[late-2000s] Video clip of Thom Yorke talking about Radiohead album  
Around the time "In Rainbows" came out, I remember seeing a video clip of Thom Yorke talking to the viewer. He did something weird and offbeat immediately, said something weird, maybe a little dance or moving his head around or something. Thanks in advance!  
↑ 2 ↓ 6

[Video] Recreation of Greys Anatomy surgery scene  
These people recreated a scene in Greys Anatomy where someone puts on the surgeons gloves for them, then they operate and the patient dies and they continue to try and revive them. The caption is something along the lines of "Greys Anatomy be like..."  
↑ 1 ↓ 1

7 comments  
Add a Comment

SassyPants5 27 • 53 Seconds later  
Poison Ivy  
↑ 7 ↓ 0 Reply

sonz82 18 • 5 Minutes later  
Wow that was fast, yes this is the one! Thank you!  
Solved!

6 comments  
Add a Comment

[deleted] 27 • 3 Days later  
Was it **this one** from "on the Big Fat Quiz of the Year"?  
↑ 1 ↓ 0 Reply

savorie 18 • 4 Days later  
Solved! OMG!! THIS WAS IT!  
how did you figure it out?

1 comments  
Add a Comment

PhDTotoro 13 • 7 Days later  
I remember this one. But I don't think it was Grey's Anatomy related.  
↑ 1 ↓ 0 Reply

not solved!

Observation:

Our assumption allows assessing the performance of a question without requiring a link to the known item

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction

- ❑ We crawl all questions and comments from [reddit.com/r/tipofmytongue/](https://www.reddit.com/r/tipofmytongue/)
- ❑ 1.28 million questions
- ❑ Crawled: 2009—2022
- ❑ 47 % are solved
- ❑ Many categories
  - Top-1 category: song with 160,000 questions
  - Top-2 category: movies with 159,000 questions
  - 13 categories with more than 10,000 questions

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Dataset construction Dataset construction

- ❑ We crawl all questions and comments from [reddit.com/r/tipofmytongue/](https://www.reddit.com/r/tipofmytongue/)
- ❑ 1.28 million questions
- ❑ Crawled: 2009—2022
- ❑ 47 % are solved
- ❑ Many categories
  - Top-1 category: song with 160,000 questions
  - Top-2 category: movies with 159,000 questions
  - 13 categories with more than 10,000 questions

Example data entry (jsonl format):

Question				Answer	
Title	Content	UTC	Solved	Content	UTC
[Movie] From the 1990s?	I remember a movie where a young woman having an affair with a married man [...]	1576446820	✓	[Poison Ivy] ( <a href="https://www.imdb.com/title/tt0105156/">https://www.imdb.com/title/tt0105156/</a> )	1576446873



# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Pilot study: Known-Item Question Performance Prediction

- We use the qpptk toolkit<sup>1</sup>
- We apply all seven pre-retrieval query performance predictors to our dataset

---

<sup>1</sup><https://github.com/Zendelo/QPP-EnhancedEval/tree/main/code/qpptk>

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Pilot study: Known-Item Question Performance Prediction

- We use the qpptk toolkit<sup>2</sup>
- We apply all seven pre-retrieval query performance predictors to our dataset

max-idf			avg-idf			scq			avg-scq			var			max-var			avg-var		
$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$
.06	.06	.07	.02	.02	.07	.11	.16	.09	-.05	-.06	.02	.12	.18	.10	.08	.10	.03	-.06	-.07	.00

<sup>2</sup><https://github.com/Zendelo/QPP-EnhancedEval/tree/main/code/qpptk>

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Pilot study: Known-Item Question Performance Prediction

- We use the qpptk toolkit<sup>3</sup>
- We apply all seven pre-retrieval query performance predictors to our dataset

max-idf			avg-idf			scq			avg-scq			var			max-var			avg-var		
$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$	$\tau$	$\rho$	$r$
.06	.06	.07	.02	.02	.07	.11	.16	.09	-.05	-.06	.02	.12	.18	.10	.08	.10	.03	-.06	-.07	.00

All predictors fail!

<sup>3</sup><https://github.com/Zendelo/QPP-EnhancedEval/tree/main/code/qpptk>

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Conclusion

- ❑ We presented a large dataset for known-item question performance prediction
- ❑ Proxy for the performance of a question = time elapsed until question is solved
- ❑ Pilot study on pre-retrieval predictors
  - None could really predict a known-item question's performance.
- ❑ Data and code available:
  - <https://github.com/webis-de/QPP-23>

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Conclusion

- ❑ We presented a large dataset for known-item question performance prediction
- ❑ Proxy for the performance of a question = time elapsed until question is solved
- ❑ Pilot study on pre-retrieval predictors
  - None could really predict a known-item question's performance.
- ❑ Data and code available:
  - <https://github.com/webis-de/QPP-23>

## Future Work

- ❑ Ensure that there is no overlap to the TREC test topics
- ❑ Simplify data wrangling via `ir_datasets`
- ❑ Apply BERT-style predictors

We hope that the dataset supports the upcoming tip-of-my-tongue known-item retrieval track at TREC 2023

# A Large-Scale Dataset for Known-Item Question Performance Prediction

## Conclusion

- ❑ We presented a large dataset for known-item question performance prediction
- ❑ Proxy for the performance of a question = time elapsed until question is solved
- ❑ Pilot study on pre-retrieval predictors
  - None could really predict a known-item question's performance.
- ❑ Data and code available:
  - <https://github.com/webis-de/QPP-23>

## Future Work

- ❑ Ensure that there is no overlap to the TREC test topics
- ❑ Simplify data wrangling via `ir_datasets`
- ❑ Apply BERT-style predictors

We hope that the dataset supports the upcoming tip-of-my-tongue known-item retrieval track at TREC 2023

**Thank You!**