NTCIR-19 Tip of the Tongue Task Introduction

Jaime Arguello
UNC

Fernando Diaz
CMU

Bhaskar Mitra

To Eun Kim
CMU

Maik Fröbe
Uni Jena

Tip of the Tongue (ToT) Known-Item Retrieval

- An **identification task** where the searcher has previously experienced or consumed an item but **cannot recall a reliable identifier**.
- Named after the psychological state of not being able to recall a specific word but being able to recall words of similar form or meaning.
- Existing search systems are ineffective in resolving ToT information needs.

Tip of the Tongue Known-Item Retrieval

A Case Study in Movie Identification

Jaime Arguello, University of North Carolina at Chapel Hill, jarguell@email.unc.edu
Adam Ferguson, Microsoft, adfergus@microsoft.com
Emery Fine, Microsoft, emfine@microsoft.com
Bhaskar Mitra, Microsoft, bhaskar.mitra@microsoft.com
Hamed Zamani*, University of Massachusetts Amherst, zamani@cs.umass.edu
Fernando Diaz†, Microsoft, diazf@acm.org

ABSTRACT

While current information retrieval systems are effective for known-item retrieval where the searcher provides a precise name or identifier for the item being sought, systems tend to be much less effective for cases where the searcher is unable to express a precise name or identifier. We refer to this as *tip of the tongue (TOT) known-item retrieval*, named after the cognitive state of not being able to retrieve an item from memory. Using movie search as a case study, we explore the characteristics of questions posed by searchers in TOT states in a community question answering website. We analyze how searchers express their information needs during TOT states in the movie domain. Specifically, what information do searchers remember about the item being sought and how do they convey this information? Our results suggest that searchers use a combination of information about: (1) the *content* of the item sought, (2) the *context* in which they previously engaged with the item, and (3) previous attempts to find the item using other resources (e.g., search engines). Additionally, searchers convey information by sometimes expressing uncertainty (i.e., hedging), opinions, emotions, and by performing relative (vs. absolute) comparisons with attributes of the item. As a result of our analysis, we believe that searchers in TOT states may require specialized query understanding methods or document representations. Finally, our preliminary retrieval experiments show the impact of each information type presented in information requests on retrieval performance.

ACM Reference Format:
Jaime Arguello, Adam Ferguson, Emery Fine, Bhaskar Mitra, Hamed Zamani, and Fernando Diaz. 2021. Tip of the Tongue Known-Item Retrieval: A Case Study in Movie Identification. In *Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '21), March 14–19, 2021, Canberra, ACT, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3406522.3446021

*A part of this work was done while Hamed Zamani was affiliated with Microsoft.
†Now at Google.

CHIIR '21, March 14–19, 2021, Canberra, ACT, Australia

ACM ISBN 978-1-4503-8055-3/21/03…$15.00
https://doi.org/10.1145/3406522.3446021

1 INTRODUCTION

Known-item retrieval refers to a broad class of scenarios where a searcher's information need is for a single, specific document known to exist in the corpus [26]. Searchers may seek an item they have seen before [10] or one they believe exists [4]. During known-item retrieval, searchers may express their need using a unique identifier (e.g., title), bibliographic information (e.g., author or genre), or content cues (e.g., keywords). Information retrieval systems, for example web search engines, can leverage previously issued queries and engagements to improve known-item search, although such techniques are less effective for unpopular or new documents with less behavioral data [9].

In this paper, we investigate known-item retrieval scenarios where a searcher is looking for a previously seen item but is *not* able to express precise or even reliable information about the item. Imprecision may result from a long delay between the information need and when the searcher most recently engaged with the item (i.e., long-term memory degradation) or from the lack of a universally adopted content description language (e.g., searching for a song based on its drum beat or a book based on its narrative structure). While similar to re-finding tasks, the emphasis on identification—as opposed to navigation—makes these information needs acute, one-off episodes—as opposed to repeated requests.

We refer to a searcher in this situation as being in a 'tip of the tongue' (TOT) state, due to its similarity to the cognitive state of not being able to retrieve an item from memory [6]. We adopt the following definition of a TOT information need,

> *an item identification task where the searcher has previously experienced or consumed the item but cannot recall a reliable identifier.*

Elsweiler et al. [12] found that searchers specifically in TOT states tend to be prone to higher levels of frustration compared to other memory lapse states.

In recent years, several community question-answering (CQA) sites have emerged to support searchers with TOT information needs.[1] These sites are tailored for searchers who want to find a specific known item (e.g., a movie, book, song, artist, video game,

[1]https://www.reddit.com/r/tipofmytongue, https://irememberthismovie.com, https://www.watzatsong.com/en, https://www.goodreads.com/group/show/185-what-s-the-name-of-that-book, https://scifi.stackexchange.com/questions/tagged/story-identification

WatZat Song?

I Remember This Movie...

Remember a scene from a movie, but not the title? Here's where you can get help, and help others recover forgotten movie memories.

goodreads Home My Books Browse ▾ Community ▾

What's the Name of That Book??? discussion

it's on the...
r/TipOfMyTongue

TOMT: When you can't remember that...thing...
r/tipofmytongue

NameThatMovie?

StackExchange [story-identification]

Science Fiction & *Fantasy*

Home

Questions tagged [story-identification]

Ask Question

more general phenomenon than movies

What do TOT Information Requests Look Like?

Example

Ok so I don't really remember anything but one scene, so I will try to give as much detail as I can. I saw this movie when I was very young and only remember this scene because (from what I remember) I don't think I actually watched most of the movie because I found it pretty scary, so it is probably not a children/family movie. I think I watched it in 2006 (mid-end of the year probably) and it was on a tv in someone's house, so it was old enough to be released on tv/dvd (not still in cinemas). It was in English and colour I'm pretty sure. So this scene: There was a (or multiple) giant robot-like things and I think they were sort of sphere shaped. It was destroying a city and going around picking people up (possible killing them?) in giant net-like things i think. There was a father and daughter (i think it was a daughter) and they ended up getting separated because the father got picked up by the robot thing. When the robot thing was picking people up there was a lot of red liquid stuff (quite possibly blood, but maybe something else?). In the end I think the father made it back to the daughter. And that is all I remember, sorry if it is vague, but really hope someone can help.

Towards Understanding Complex Known-Item Requests on Reddit

Florian Meier
Aalborg University Copenhagen
Copenhagen, Denmark
fmeier@hum.aau.dk

Toine Bogers
Aalborg University Copenhagen
Copenhagen, Denmark
toine@hum.aau.dk

Maria Gäde
Humboldt-Universität zu Berlin
Berlin, Germany
maria.gaede@ibi.hu-berlin.de

Line Ebdrup Thomsen
Aalborg University Copenhagen
Copenhagen, Denmark
let@hum.aau.dk

You Know What I'm Saying: Designing Conversational Strategies of AI Agent for Tip of the Tongue Phenomenon

Juhye Ha
Graduate School of Information,
Yonsei University
Seoul, Republic of Korea
juhye0329@yonsei.ac.kr

Dayoung Lee
Graduate School of Information,
Yonsei University
Seoul, Republic of Korea
gsi.dayounglee@yonsei.ac.kr

Changhoon Oh
Graduate School of Information,
Yonsei University
Seoul, Republic of Korea
changhoonoh@yonsei.ac.kr

"Kinda like The Sims... But with ghosts?": A Qualitative Analysis of Video Game Re-finding Requests on Reddit

Ida Kathrine Hammeleff Jørgensen
Center for Computer Games Research
IT University of Copenhagen
Copenhagen, Denmark
ihjo@itu.dk

Toine Bogers
Department of Communication & Psychology
Aalborg University Copenhagen
Copenhagen, Denmark
toine@hum.aau.dk

Understanding Feeling-of-Knowing in Information Search: An EEG Study

DOMINIKA MICHALKOVA, University of Strathclyde, UK
MARIO PARRA RODRIGUEZ, University of Strathclyde, UK
YASHAR MOSHFEGHI, University of Strathclyde, UK

Browsing Lost Unformed Recollections: A Benchmark for Tip-of-the-Tongue Search and Reasoning

Sky CH-Wang[•*] **Darshan Deshpande**[◦] **Smaranda Muresan**[•]
Anand Kannappan[◦] **Rebecca Qian**[◦]
[◦]Patronus AI [•]Columbia University
{skywang, smara}@cs.columbia.edu, {darshan, anand, rebecca}@patronus.ai

When the Music Stops: Tip-of-the-Tongue Retrieval for Music

Samarth Bhargav[*]
University of Amsterdam
Amsterdam, Netherlands
s.bhargav@uva.nl

Anne Schuth
Spotify
Amsterdam, Netherlands
aschuth@spotify.com

Claudia Hauff
Spotify
Amsterdam, Netherlands
claudiah@spotify.com

On Interfacing Tip-of-the-tongue References In Movie Cases

Jongho Kim
Interdisciplinary Program in Artificial Intelligence, Seoul National University / Korea
jongh97@snu.ac.kr

Soona Hong
Seoul National University / Korea
hongsoona@snu.ac.kr

Seung-won Hwang[*]
Seoul National University / Korea
seungwonh@snu.ac.kr

Tip of the Tongue Known-Item Retrieval

A Case Study in Movie Identification

Jaime Arguello
University of North Carolina at Chapel Hill
jarguell@email.unc.edu

Adam Ferguson
Microsoft
adfergus@microsoft.com

Emery Fine
Microsoft
emfine@microsoft.com

Bhaskar Mitra
Microsoft
bhaskar.mitra@microsoft.com

Hamed Zamani[*]
University of Massachusetts Amherst
zamani@cs.umass.edu

Fernando Diaz[†]
Microsoft
diazf@acm.org

Tip of the Tongue Query Elicitation for Simulated Evaluation

Yifan He[*]
Carnegie Mellon University
Pittsburgh, PA, USA
yifanhe@alumni.cmu.edu

To Eun Kim[*]
Carnegie Mellon University
Pittsburgh, PA, USA
toeunk@cs.cmu.edu

Fernando Diaz
Carnegie Mellon University
Pittsburgh, PA, USA
diazf@acm.org

Jaime Arguello
UNC Chapel Hill
Chapel Hill, NC, USA
jarguell@email.unc.edu

Bhaskar Mitra
Microsoft Research
Montréal, QC, Canada
bmitra@microsoft.com

'It's on the tip of my tongue': A new Dataset for Known-Item Retrieval

Samarth Bhargav
s.bhargav@uva.nl
IRLab, University of Amsterdam
The Netherlands

Georgios Sidiropoulos
g.sidiropoulos@uva.nl
IRLab, University of Amsterdam
The Netherlands

Evangelos Kanoulas
e.kanoulas@uva.nl
IRLab, University of Amsterdam
The Netherlands

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

Maik Fröbe[1,*], Eric Oliver Schmidt[2] and Matthias Hagen[1]

[1]*Friedrich-Schiller-Universität Jena, 07743 Jena, Germany*
[2]*Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany*

Comparison of Different Lexical Resources With Respect to the *Tip-of-the-Tongue* Problem

Michael Zock[1], Chris Biemann[2]

A Corpus of Realistic Known-Item Topics With Associated Web Pages in the ClueWeb09

Matthias Hagen, Daniel Wägner, and Benno Stein

Bauhaus-Universität Weimar
<first name>.<last name>@uni-weimar.de

Decomposing Complex Queries for Tip-of-the-tongue Retrieval

Kevin Lin[♠] **Kyle Lo**[♡] **Joseph E. Gonzalez**[♠] **Dan Klein**[♠]
[♠]University of California Berkeley [♡]Allen Institute for AI
{k-lin,jegonzal,klein}@berkeley.edu kylel@allenai.org

Drivers of Information Needs: A Behavioural Study

Exploring Searcher's Feeling-of-Knowing

Dominika Michalkova
University of Strathclyde
Glasgow, UK
dominika.michalkova@strath.ac.uk

Mario Parra Rodriguez
University of Strathclyde
Glasgow, UK
mario.parra-rodriguez@strath.ac.uk

Yashar Moshfeghi
University of Strathclyde
Glasgow, UK
yashar.moshfeghi@strath.ac.uk

Generalizable Tip-of-the-Tongue Retrieval with LLM Re-ranking

Luís Borges
lborges@andrew.cmu.edu
Instituto Superior Técnico and INESC-ID
Lisbon, Portugal

Rohan Jha
rjha@utexas.edu
The University of Texas at Austin
Austin, Texas, United States

Jamie Callan
callan@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, United States

Bruno Martins
bruno.g.martins@tecnico.ulisboa.pt
Instituto Superior Técnico and INESC-ID
Lisbon, Portugal

Challenges

- Ignoring social niceties
- Leveraging contextual (episodic) memories
- Leveraging semantic memories about: metadata
- Leveraging semantic memories about: scenes
- Leveraging semantic memories about: plot
- Modeling levels of (un)certainty
- Modeling exclusion criteria
- Relative movie comparisons (multi-hop reasoning)
- Relative attribute comparisons (multi-hop reasoning)
- Dealing with false memories
- Not all items are equally "forgettable"

TREC Tip of the Tongue

trec-tot.github.io

TREC ToT 2023-2025

Experimental Setup

- **Protocol:** single-turn, *ad hoc* retrieval.
- **Queries:**
  - 2023: queries from the Microsoft Tip of the Tongue (MS-ToT) dataset.
  - 2024: queries from from MS-ToT augmented with simulated queries.
  - 2025: queries from from MS-ToT augmented with simulated queries.
- **Corpus:** Subset of Wikipedia, corresponding to query classes.
- **Evaluation:** NDCG@10, NDCG@1000, MRR@1000, Recall@1000. There is *one* relevant document per query.

Simulated queries

TREC Tip of the Tongue

trec-tot.github.io

1. design a prompt to generate a ToT queries by iterating on language to
    a. generate system orderings similar to system orderings generated by human ToT queries, and
    b. be linguistically similar to human ToT queries.
2. generate new ToT queries,
    a. sample a target concept associated with a random wikipedia page (relevant document)
    b. generate a ToT query based on the prompt and sampled page (query text)

Tip of the Tongue Query Elicitation for Simulated Evaluation

Yifan He*
Carnegie Mellon University
Pittsburgh, PA, USA
yifanhe@alumni.cmu.edu

To Eun Kim*
Carnegie Mellon University
Pittsburgh, PA, USA
toeunk@cs.cmu.edu

Fernando Diaz
Carnegie Mellon University
Pittsburgh, PA, USA
diazf@acm.org

Jaime Arguello
UNC Chapel Hill
Chapel Hill, NC, USA
jarguell@email.unc.edu

Bhaskar Mitra
Microsoft Research
Montréal, QC, Canada
bmitra@microsoft.com

Abstract

Tip-of-the-tongue (TOT) search occurs when a user struggles to recall a specific identifier, such as a document title. While common, existing search systems often fail to effectively support TOT scenarios. Research on TOT retrieval is further constrained by the challenge of collecting queries, as current approaches rely heavily on community question-answering (CQA) websites, leading to labor-intensive evaluation and domain bias. To overcome these limitations, we introduce two methods for eliciting TOT queries—leveraging large language models (LLMs) and human participants—to facilitate simulated evaluations of TOT retrieval systems. Our LLM-based TOT user simulator generates synthetic TOT queries at scale, achieving high correlations with how CQA-based TOT queries rank TOT retrieval systems when tested in the Movie domain. Additionally, these synthetic queries exhibit high linguistic similarity to CQA-derived queries. For human-elicited queries, we developed an interface that uses visual stimuli to place participants in a TOT state, enabling the collection of natural queries. In the Movie domain, system rank correlation and linguistic similarity analyses confirm that human-elicited queries are both effective and closely resemble CQA-based queries. These approaches reduce reliance on CQA-based data collection while expanding coverage to underrepresented domains, such as Landmark and Person. LLM-elicited queries for the Movie, Landmark, and Person domains have been released as test queries in the TREC 2024 TOT track, with human-elicited queries scheduled for inclusion in the TREC 2025 TOT track. Additionally, we provide source code for synthetic query generation and the human query collection interface, along with curated visual stimuli used for eliciting TOT queries.

CCS Concepts

• Information systems → Test collections; Specialized information retrieval; • Human-centered computing → Interaction design.

*Equal Contribution.

This work is licensed under a Creative Commons Attribution International 4.0 License.

SIGIR '25, Padua, Italy
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3730335

Keywords

Tip-of-the-Tongue Known-Item Retrieval; Synthetic Query Generation; Human Query Elicitation

ACM Reference Format:
Yifan He, To Eun Kim, Fernando Diaz, Jaime Arguello, and Bhaskar Mitra. 2025. Tip of the Tongue Query Elicitation for Simulated Evaluation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3726302.3730335

1 Introduction

Tip-of-the-Tongue (TOT) retrieval is a known-item search task where the searcher aims to re-find a previously encountered item but struggles to formalize the query due to an inability to recall specific identifiers, such as a document title or a person's name [4, 10]. Unlike other known-item retrieval tasks, TOT information requests are typically verbose and complex, often blending semantic details about the context in which the searcher previously engaged with the item [25, 27]. These queries frequently include linguistic phenomena such as uncertainty, exclusion criteria, relative comparisons, and even false memories [4, 27], making them especially challenging for traditional retrieval systems [2, 4, 10, 26].

Addressing the challenges of TOT retrieval is critical because searchers in the TOT state experience higher level of frustration compared to other types of memory-related search failures [18]. TOT retrieval spans a wide range of domains, from casual leisure search [19], such as finding games [23], music [9], and books [10], to critical applications like email retrieval [16, 17, 24] and enterprise content search [15]. In professional settings, ineffective TOT support can lead to significant economic costs, as employees spend excessive time retrieving essential information, reducing productivity and efficiency [37].

The evaluation of TOT retrieval systems faces significant challenges due to limitations in test collection resources and current dataset construction methods. Most existing TOT queries in current datasets are sourced from community question answering (CQA) websites, which fail to fully capture the diversity of real-world TOT information needs [27]. This reliance on CQA platforms introduces domain skewness, as existing datasets and studies are heavily biased toward casual leisure topics like movies and books [11, 19], while underrepresenting areas such as people and landmarks. Even attempts to expand domain coverage [27] have had limited success,

3398

NTCIR-19 ToT

What do TOT Information Requests Look Like?

Example

Ok so I don’t really remember anything but one scene, so I will try to give as much detail as I can. I saw this movie when I was very young and only remember this scene because (from what I remember) I don’t think I actually watched most of the movie because I found it pretty scary, so it is probably not a children/family movie. I think I watched it in 2006 (mid-end of the year probably) and it was on a tv in someone’s house, so it was old enough to be released on tv/dvd (not still in cinemas). It was in English and colour I’m pretty sure. So this scene: There was a (or multiple) giant robot-like things and I think they were sort of sphere shaped. It was destroying a city and going around picking people up (possible killing them?) in giant net-like things i think. There was a father and daughter (i think it was a daughter) and they ended up getting separated because the father got picked up by the robot thing. When the robot thing was picking people up there was a lot of red liquid stuff (quite possibly blood, but maybe something else?). In the end I think the father made it back to the daughter. And that is all I remember, sorry if it is vague, but really hope someone can help.

What do TOT Information Requests Look Like?

Example

航空运输人员被困荒岛，独自与排球朋友 维森 活数年，决定离开孤岛，最终回到以前的生活！电影讲述了孤自一人如何求生过程，非常经典，这部外国电影叫什么啊！

映画のタイトルが思い出せないので、教えてください。80年代頃の、日本の高校、恋愛、男の子がバスケ部で、朝の新聞配達をしている、女の子の家が教会、バスケの試合途中で映画終了、こんな感じです。当時のアイドル人気の為に作ったような、たいした内容ではない、当時の粗雑乱造映画のような印象があります。ご存知の方がいらっしゃいましたら、回答よろしくお願いいたします。

제가 몇년전에 봤던 애니메이션이 지금 너무너무 보고싶은데 제목이 기억이안나서 질문 드립니다.. 기억나는걸 적어보자면 한편으로된 영화였고 여름이 배경이였고 여주 남주 둘다 쇼핑몰 같은곳에서 일하는 설정이였어요 어르신분들 케어해주는곳? 에서 도와드리고 남주였는지 여주였는지 기억이안나는데 옥상에서 스케이트보드도타고 노래듣는 엠피쓰리? 같은거 떨어트려서 주워주다가 눈마주침하는 장면도 있었고 끝날때쯤 불꽃놀이 보러간 장면이 있었는데 남주가 엄청 뛰어서 보러가고 클라이맥스되는 장면도 있었던거 같습니다.. 제목 좀 꼭 좀 부탁드려요

TREC Limitations

- **queries were restricted to English language**
    - **NTCIR RQ: How does performance vary across non-English languages?**
- **domains were predominantly US-based**
    - **NTCIR RQ: How does performance vary across east Asian domains/topics?**
        - synthesize topics that are, based on Wikipedia page views,
            i. **more** represented in the target language than other languages,
            ii. **equally** represented in the target language and other languages, and
            iii. **less** represented in the target language than other languages.

Summary

- We look forward to your participation in NTCIR-ToT!

Addendum

Example Prompt Template for Landmark Queries

Let's do a role play. You are now a person who vaguely remembers a place called {ToTObject}. You are trying to recall the name of the place by posting a verbose post in an online forum like Reddit describing the place. Generate a post of around 200 words about the place {ToTObject}. Your post must describe a vague memory of the place without revealing its exact name. People on the forum must have a hard time figuring out which place you are looking for. The answer should be difficult to find in search engines, so avoid using obvious keywords. I will provide you with some basic information about the place, and you must follow the guidelines to create a post.

Information about {ToT target name}: {first paragraph of Wikipedia article}

Guidelines:
MUST FOLLOW:

[1] Reflect the imperfect nature of memory with phrases that express doubt or mixed recollections, avoiding direct phrases like "I'm not sure if it is true, but".

[2] Do not directly specify the name of the place.

[3] Refer to the places in an ambiguous way using descriptions instead of names.

[4] Maintain a casual and conversational tone throughout the post, making sure it sounds natural and engaging without using formal structures.

[5] Provide vivid but ambiguous details to stir the reader's imagination while keeping them guessing.

[6] Use the provided information only as inspiration to craft a unique and engaging narrative, avoiding any direct replication of the given phrases.

[7] Start directly with your post, avoiding formal greetings like "Hello" or "Hey everyone."

[8] Start directly with your post, without describing your state of mind like "So, there's this", "I remember", "I've been thinking about".

COULD FOLLOW:

[1] Share a personal anecdote about your time at the place and the people you were with, but avoid common phrases like "When I was young." Instead, find unique ways to set the scene.

[2] Focus on sensory details like the overall mood, sounds, and emotional impact of being in the place, using vivid descriptions.

[3] Draw comparisons with other places or familiar experiences in a nuanced way that doesn't directly echo well-known locations.

[4] Introduce a few incorrect or mixed-up details to make the recollection seem more realistic and harder to pinpoint.

[5] Describe particular scenes or moments using ambiguous terms or partial descriptions.

[6] End the post by encouraging responses with genuine, open-ended questions for help.

Generate a post based on these guidelines.