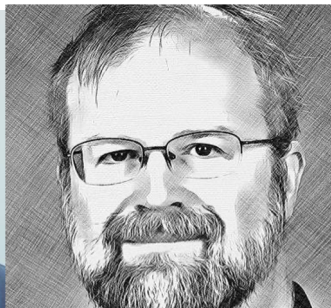


TREC RAG 2024



What are all the tracks in TREC 2023?

Sources
Call for New TREC tracks
nist · 1

Call to TREC 2023 - Text Retrieval Conference...
nist · 2

TREC 2023 Deep Learning Track...
microsoft · 3

Answer

- The TREC 2023 (Text Retrieval Conference) consists of several a specific facet of the retrieval problem. The tracks for TREC 20
1. Deep Learning Track: This track includes passage ranki ranking tasks, with two subtasks in each case: full rank
 2. ATOMIC (Authoring Tools for Multimedia Content) Tra authoring tools for multimedia content and includes requirements, and evaluation methods
 3. Product Search Track: This is the first year of the which focuses on the creation of a reusable colle
 4. Other Tracks: TREC 2023 also contains tracks e Trials, CrisisFACTS, and NeuCLIR, which ran in
- These tracks aim to encourage research in informat applications by providing a large test collection, un forum for organizations interested in comparing th tasks for each track are formulat on the sponsors, as well as the suitabil

ChatGPT 4

GPT-4
With DALL-E, browsing and analysis

Tell me about the nobel Prize laureates in physics 2023

The Nobel Prize in Physics 2023 was awarde experimental methods that generate attose



Pierre Agostini, Nobel Prize laure...

The laureates, who worked independe durations as short as a few hundred at the pulses used in previous laser rese

Attosecond pulses are so short that t them a powerful tool for studying ch

The laureates' work has led to a new possibilities for developing new tec

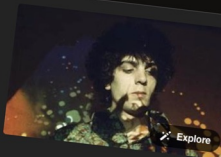
Agostini, Krausz, and L'Huillier's w committee said that their discover have opened up new possibilities

- Sources
1. www.nobelprize.org/#:~:text=T

Microsoft Bing
SEARCH CHAT View other drafts
Creative Balanced Precise

- ✓ Searching for: Syd Barrett
- ✓ Generating answers for you...

Syd Barrett was an English singer, guitarist and songwriter who co-founded the rock band Pink Floyd in 1965. He was accented singing, and stream-of-consciousness writing style. He was influential for his free-form playing and for employing effects such as dissonance, distortion, echo and feedback



He led Pink Floyd as the frontman and primary songwriter for their first album The Piper at the Gates of Dawn (1967), portions of their second album A Saucerful of Secrets (1968), and several songs that were not released until years later. However, he was ousted from the band in 1968 due to his erratic behavior and mental health issues, which were worsened by his use of psychedelic drugs

He then pursued a brief solo career, releasing two albums The Madcap Laughs (1970) and Barrett (1970), with the help of some of his former bandmates. He left the music industry in 1972 and retired from public life. He continued painting and gardening until his death from pancreatic cancer in 2006

He is regarded as a cult figure and a source of inspiration for many musicians and artists. Pink Floyd recorded several tributes to him, such as the song suite "Shine On You Crazy Diamond" and parts of the rock opera The Wall. He was inducted into the Rock and Roll Hall of Fame as a member of Pink Floyd in 1996

If you want to learn more about Syd Barrett, you can visit his official website or check out his discography

Learn more: 1. en.wikipedia.org 2. theguardian.com 3. sydbarrett.com +2 more

Who is Syd Barrett?

**All these systems and hype,
but *where* is the *evaluation*?**

Introduction

We believe there are two tasks at the core of any RAG system:

Retrieval Task: Given a query, retrieve a ranked list of passages.

Retrieval-Augmented Generation Task: Generate answers that include attributions to the specific supporting passages in the ranked list that *support* (ground) the generation.

Participants submit well-attributed RAG answers along with their supporting passage list!

Evaluation Workflow (*Proposed*)

1. Gather answers from participants
2. Pool *cited* reference passages
3. Evaluate *relevance* of pooled passages
4. Evaluate *support* of cited references – Attribution Check
5. *Nuggetize* the pool of relevant passages
6. *Assign* nuggets to participant answers

Yes, all these seem similar to tracks from prior years, but that's a win, more learnings to guide our decisions!

Retrieval Task

Collection: Given the three years of relevance judgments and data, that comprise the *MS MARCO v2 Passage Ranking collection* (and methods spanning the years before that too with the v1 collection), we believe it is apt choice for the Y1 of our track!

Queries: We aim to curate some 50 *difficult* topics capturing *long-form, definition, list, ambiguous* information needs. These would likely come from a larger set of ~200 topics that have been *filtered* if *easy* for an LLM in isolation or fully captured in a single passage.

Sample Information Need/Answer

Tell me about the Cassini space probe.

Launched in 1997, the Cassini-Huygens mission, a collaboration between NASA, ESA, and ASI, was the first to enter Saturn's orbit and the fourth to study the planet [1][10]. This ambitious mission included the Cassini orbiter and the Huygens probe, which landed on Titan, Saturn's largest moon, in 2005 [2][3][6]. Cassini's arrival at Saturn in 2004 marked the beginning of an extensive exploration, revealing new complexities in Saturn's moons and rings [5][7]. After a groundbreaking two-decade journey, the mission concluded in 2017 [8].

[1] reference to an MS MARCO v2 passage

[2] reference to an MS MARCO v2 passage

[3] ...

Evaluation Workflow – Retrieval

:

Step 2: Pool all the cited passages.

Step 3: Evaluate passages for relevance

Step 4: Go back to answer: does the cited passage actually support the sentence?

:

NIST Assessors can do their magic! Step 3 likely similar to TREC DL '23.

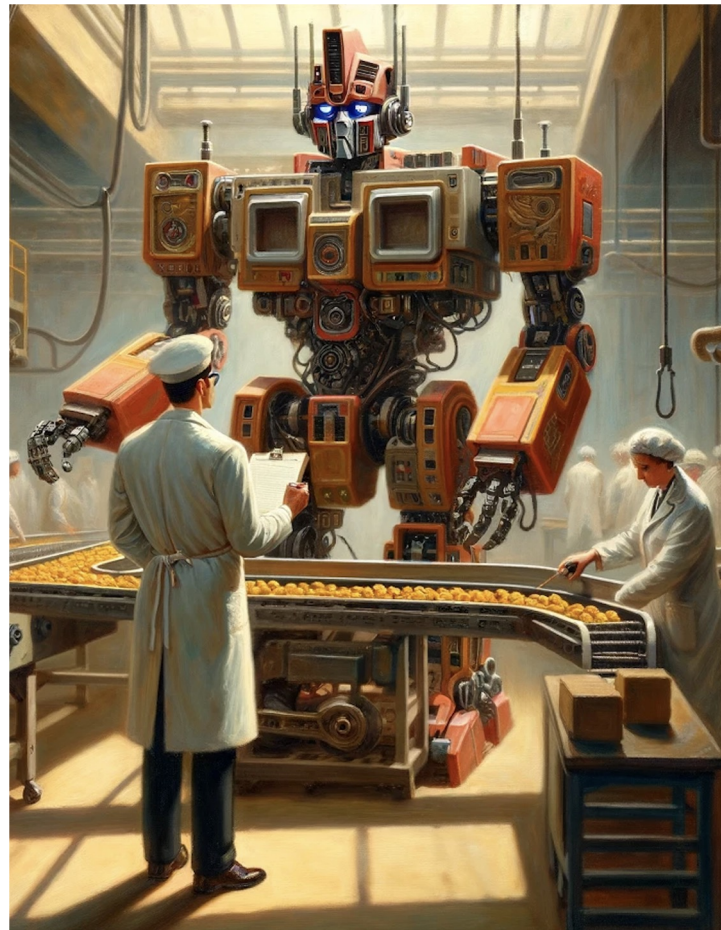
Step 5: *Nuggetize*

In prior TREC QA tracks, these were done entirely by a NIST assessor.

Here, we envision an interactive human-in-the-loop setup.

LLM peruses all the pooled passages to provide *atomic nuggets ordered by importance* and the NIST annotator can *modify/remove/add* to the ordering.

We hope this can transition into a self-iterative LLM-as-a-nuggetizer system for evaluation.



After some GPT₄ prompt engineering...

What is the Cassini space probe?

Collaboration between NASA, ESA, and ISA

Sent to study Saturn and its moons

First probe to enter Saturn's orbit

4th space probe to study Saturn

Comprised NASA's Cassini probe, ESA's Huygens lander

Huygens designed to land on Titan

Launched in 1997

Took first close photos of Saturn's rings

...

Named after astronomers Cassini, Huygens

1	<i>vital</i>	32 kilograms plutonium powered
2	<i>vital</i>	seven year journey
3	<i>vital</i>	Titan 4-B Rocket
4	<i>vital</i>	send Huygens to probe atmosphere of Titan, Saturn's largest moon
5	<i>okay</i>	parachute instruments to planet's surface
6	<i>okay</i>	oceans of ethane or other hydrocarbons, frozen methane or water
7	<i>vital</i>	carries 12 packages scientific instruments and a probe
8	<i>okay</i>	NASA primary responsible for Cassini orbiter
9	<i>vital</i>	explore remote planet and its rings and moons, Saturn
10	<i>okay</i>	European Space Agency ESA responsible for Huygens probe
11	<i>okay</i>	controversy, protest, launch failure, re-entry, lethal risk, humans, plutonium
12	<i>okay</i>	Radioisotope Thermoelectric Generators, RTG
13	<i>vital</i>	Cassini, NASA'S Biggest and most complex interplanetary probe
14	<i>okay</i>	find information on solar system formation
15	<i>okay</i>	Cassini Joint Project between NASA, ESA, and ASI (Italian Space Agency)
16	<i>vital</i>	four year study mission

Used in mergers and acquisitions

Known for "Appalachian Spring"

USER

You need to provide atomic nuggets of information (1-9 words) that can best provide the question leveraging only the contexts provided below. Generate all your nuggets in a Pythonic list format. Make sure there is no redundant information. Order them in decreasing order of importance. Prefer nuggets that provide more interesting information.

Question: What is the Cassini space probe?

Contexts:

[1] Cassini – 1997. Cassini-Huygens, more commonly known as Cassini, was a collaboration between NASA, the European Space Agency (ESA), and the Italian Space Agency (ISA), to send a probe to study Saturn, all of its moons, and the famous rings. Launched in 1997, Cassini was the first probe to enter Saturn's orbit, and the 4th space probe to study Saturn.

[2] Cassini was a collaboration between the United States' NASA, the European Space Agency ("ESA"), and the Italian Space Agency ("ASI") to send a probe to study the Saturnian system, including the planet, its rings, and its natural satellites. The Flagship-class uncrewed robotic spacecraft comprised both NASA's Cassini probe, and ESA's Huygens lander which was designed to land on Saturn's largest moon, Titan.

[3] About the Mission. Cassini was one of the most ambitious efforts ever mounted in planetary exploration. A joint endeavor of NASA, ESA (the European Space Agency) and the Italian space agency (ASI), Cassini was a sophisticated robotic spacecraft sent to study Saturn and its complex system of rings and moons in unprecedented detail. Cassini carried a probe called Huygens to the Saturn system.

[4] But, in 1997, the Cassini-Huygens space craft was launched to explore Saturn and its moons. The space craft consists of two parts: the Cassini probe, which is an orbiter and is still active, and the Huygens probe, which was an atmospheric entry probe designed to enter the atmosphere of Titan and land on its surface.

[5] Cassini Orbiter. Huygen's Probe. Navigation. Cassini Orbiter Huygen's Probe Navigation. The Cassini-Huygens spacecraft during vibration and thermal testing in 1996. Cassini-Huygens was one of the most ambitious missions ever launched into space.

[6] Space - Cassini-Huygens Probe. 1. The Cassini-Huygens orbiter and probe entered Saturn's orbit and took the first close photos of Saturn's rings during July, 2004. 2. The space craft was launched in October of 1997 and made a seven year journey over 2.2 billion miles. 3.

[7] Cassini Huygens is a sophisticated and one of the largest interplanetary spacecraft sent to Saturn to study the ringed planet in detail over a four-year period as part of a joint mission of the NASA, ESA and the Italian space agency ASI. Giovanni Domenico Cassini. © NASA. The ESA Huygens is a scientific probe that was released by way of a parachute from the main spacecraft through the atmosphere to land on Saturn's largest moon, Titan.

[8] Cassini spacecraft at Titan. In 2017, the Cassini spacecraft wrapped up its two-decade-long mission to Saturn. Launched on October 15, 1997, the space craft arrived at Saturn on June 30, 2004. Upon its arrival, Cassini dropped the Huygens probe built by the European Space Agency.

[9] Cassini spacecraft at Titan. In 2017, the Cassini spacecraft wrapped up its two-decade-long mission to Saturn. Launched on October 15, 1997, the space craft arrived at Saturn on June 30, 2004. Upon its arrival, Cassini dropped the Huygens probe built by the European Space Agency. Huygens was equipped to study Titan by landing on the Saturn moon and achieved astounding results.

[10] Cassini was the fourth space probe to visit Saturn and the first to enter its orbit. The craft were named after astronomers Giovanni Cassini and Christiaan Huygens. Launched aboard a Titan IVB/Centaur on 15 October 1997, Cassini had a nearly 20-year life span, with 13 of these years spent orbiting Saturn.

Answer:

ASSISTANT

['Collaboration between NASA, ESA, and ISA', 'Sent to study Saturn and its moons', 'First probe to enter Saturn's orbit', '4th space probe to study Saturn', 'Comprised NASA's Cassini probe, ESA's Huygens lander', 'Huygens designed to land on Titan', 'Launched in 1997', 'Took first close photos of Saturn's rings', 'One of the largest interplanetary spacecraft', 'Mission ended in 2017', 'Named after astronomers Giovanni Cassini and Christiaan Huygens']

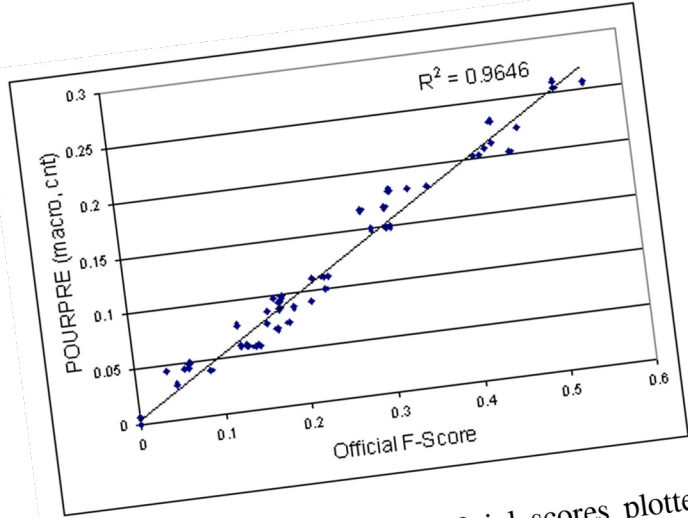


Figure 3: Scatter graph of official scores plotted against the POURPRE scores (macro, count) for TREC 2003 ($\beta = 5$).

ment

n 1997.

, and ASI.

orbit.

an in 2005.



– does the answer contain the given nuggets?

An LLM can probably achieve a more-refined *POURPRE* (Lin and Demner-Fushman, 2005) where they match nuggets by summing the unigram co-occurrences between terms from each nugget and terms from the system response.

USER

You need to score the provided answer (Answer:) for each of the nuggets following the schema:

0 - if the nugget information is missing in the answer,

1 - if the nugget information is only partially accounted for in the text

2 - if the nugget information is properly accounted for in the text

Return a Pythonic list of scores (Scores:), ensuring it is ordered by the nugget order.

Question: What is the Cassini space probe?

Answer: Launched in 1997, the Cassini-Huygens mission, a collaboration between NASA, and ASI, was the first to enter Saturn's orbit and the fourth to study the planet [1][10]. This ambitious mission included the Cassini orbiter and the Huygens probe, which landed on Titan, Saturn's largest moon, in 2005 [2][3][6]. Cassini's arrival at Saturn in 2004 marked the beginning of an extensive exploration, revealing new complexities in Saturn's moons and rings [5][7]. After a groundbreaking two-decade journey, the mission concluded in 2017 [8].

Nuggets: ['Collaboration between NASA, ESA, and ISA', 'Sent to study Saturn and its moons', 'First probe to enter Saturn's orbit', '4th space probe to study Saturn', 'Comprised NASA's Cassini probe, ESA's Huygens lander', 'Huygens designed to land on Titan', 'Launched in 1997', 'Took first close photos of Saturn's rings', 'One of the largest interplanetary spacecraft', 'Mission ended in 2017', 'Named after astronomers Giovanni Cassini and Christiaan Huygens']

Scores:

ASSISTANT

[1, 2, 2, 2, 2, 2, 0, 0, 2, 0]

Additional Evaluation (NIST Assessors/LLM-Powered)

1. Fluency/Articulation – *Assess the clarity and coherence of the response*
2. Hallucinations – *Does this answer include hallucinations?*
3. RAG as a Passage? – *How would the RAG answer stand compared to corpora passages in terms of relevance?*
4. Pairwise Evaluation – *Which of these RAG answers are better on criteria X?*

Y2 – How do we deal with stale facts? What about conflicting evidence beyond that? Sources with Misinformation?

All conditioned on \$\$\$s

That's all from me!
Questions/Feedback?

TREC RAG 2024

Ronak Pradeep, Nandan Thakur, Ehsan Kamaloo, Jimmy Lin, Nick Craswell, Ian Soboroff

Overview

Two Tasks: 1) *Retrieval Task*: Rank passages for given queries.

2) *Retrieval-Augmented Generation (RAG) Task*: Generate answers with supporting passage attributions.

Aim: Enhance retrieval and generation effectiveness, focus on varied information needs in an evolving world.

Data Source: *MS MARCO v2 Passage Ranking*. Focus: Curate ~50 topics that capture long-form, definition, list, and ambiguous information needs.

Evaluation Process: Submit RAG answers and passage lists, assess *relevance*, *supportiveness* and RAG eval following *LLM-in-the-loop nuggetization and assignment*. Additional evals considering fluency, hallucinations, and pairwise comparisons.