TREC RAG
Task 2: Support Assessment

Version 2.0, September 21, 2025, Ian Soboroff

Overview

For the RAG track, computer systems have been given a set of topics (where each topic is a question with a complex answer, such as "what is vicarious trauma and how can it be coped with"), and they have retrieved text passages and have generated answers in response to each question, including citation of which passages support which sentences in the response.

The mission of RAG assessors will be to assess the quality of both the retrieved passages and the generated responses. You will perform 3 types of assessments for each assessed topic:
1. Passage Relevance: Judge whether the retrieved passage provides an answer to the question.
2. **Support: Judge whether the generated response is supported by the passages that are cited as providing support for the response.**
3. Nuggets: Create a list of information nuggets (small "facts") that should be returned in a good answer to the question, and determine which nuggets are included in each generated response.

In RAG assessment Task 1, you have already judged the quality of the retrieved passages for some topics.

For this task, RAG assessment Task 2, you will judge whether the generated response is supported by the passages that are cited as providing support for the response. The AI systems that are generating these responses can sometimes generate sentences that make statements that can't be backed up by any sources; therefore, it's important that these AI systems be able to cite sources (Web passages) that support the statements made in their sentences.

The first thing you will do for Task 2 is select topics to work on, from a predefined list of topics listed in this Google Document for "RAG topics":
https://docs.google.com/spreadsheets/d/1ROJqHnRZEVPuO7KUczy4ytqJ86PH6oZl4fxUb_DXdC0/edit?usp=sharing

In the Google Document that has been shared with you, there are 105 candidate queries for Task 2. In the first column, I have put your name if the topic is fully judged in Task 1.

To start a topic, scroll through the candidates and select a query that has your name in the "Assessor" column.

You'll then judge Support for responses that were generated for that query. When you're done judging Support for a query, put an 'X' in the "DONE" column and pick another candidate query to judge. Continue with Support assessment for additional queries until you've used up your remaining hours for Task 2 or the remaining hours on your contract (whichever comes first).

==The remainder of this document describes assessment Task 2: Support assessment. We will have only a single phase for Support assessment for each topic. You will not judge whether a sentence requires a citation. Instead, you will only be shown pairs of cited passages and sentences and will judge whether the cited passage supports the sentence.==

Tour

Begin assessing a topic by going to the following Google Drive folder for this task:

https://drive.google.com/drive/folders/1hDuEM-swMcVTVOGmfTIeEgnpBSGG4SW1?usp=drive_link

Search for your topic in the list and double click on your topic to open a new browser tab or window that has a Google spreadsheet where you can type in your judgments for your topic. An example Google spreadsheet looks like this:

The Google spreadsheet has the following columns:

- **Cited Passage**: the passage that was cited for this Sentence
- **Sentence**: the sentence that has cited the passage in Cited Passage
- 3 empty Columns for possible judgment values for how much support the **Cited Passage** provides for the **Sentence**:
    - **Full**
    - **Partial**
    - **None**
- **Sentence Context**: The full response in which the sentence appears; the sentence itself is displayed in **bold** typeface in the Sentence Context. Because sentences in the Sentence column are not displayed in the order in which they appear in a response (and in fact, consecutive sentences in the **Sentence** column may come from different responses) the Sentence Context column can help you interpret the sentence, in case you need more context to understand what the sentence is saying (e.g., if the sentence has a pronoun and you need to determine what the pronoun is referring to).
- **Completed?**: A value that is either "Yes" or "No", and that is filled in automatically by the spreadsheet to show whether you have completed assessment for this row. The value will be "Yes" only if you have put an "x" under exactly one of the 3 possible judgment values (Full, Partial, or None) for the row.

The only cells that you should modify in the spreadsheet are the cells in the "Full", "Partial", and "None" Columns (Columns C, D, and E). Do not add/delete/modify the text in any other cells under any other columns, as that will mess up the spreadsheet.

For each row, put an "x" under *one* of the possible judgment values (Full, Partial, or None) based on your assessment of the amount of support that the current passage ("Cited Passage" in Column A) provides for the current Sentence (in Column B). If you change your mind about a judgment that you have made, you'll have to delete the "x" for the old judgment (Full, Partial, or None) and put an "x" under the correct column for the new judgment (Full, Partial, or None).

The spreadsheet will not display the sentences in the order in which they appear in a response; instead, the spreadsheet has grouped together all of the sentences (across all responses for the topic) that cite the same passage, and displays them consecutively under the Sentence column; thus, you would be able to read each passage just once per topic, and then step through all of the sentences that cite that passage. (If you need to see more context to interpret the sentence, you can look at the Sentence Context column to view the whole response in which the sentence appears; but for the most part, you should be able to ignore the Sentence Context column.)

The text of the Cited Passage is displayed once for each sentence that cites it. Consecutive passages in the Cited Passage column will be identical *until* you encounter a Cited Passage that is highlighted in yellow; this indicates that the yellow Cited Passage is a new passage (and subsequent Cited Passages with be the same as this yellow Cited Passage, until you encounter

In order to view only the unjudged rows, you should click on "Create a Filter" (the funnel icon in the menu of the Google sheet). Next (after the funnel icon has been selected), click on the small button beside "COMPLETED?" and add and remove checkmarks next to the values for "COMPLETED" to show only rows that match that value; for example, to show rows that are not completed, ensure that only "No" has a checkmark next to it, and then click the green "OK" button to apply the filter.

Support Annotation

For Support Annotation, you will be shown a single cited passage for each sentence and will evaluate whether the sentence is supported by its corresponding cited passage. You should interpret the sentence in the context of the response in which it appears (i.e., the "Sentence Context" of the Google spreadsheet). Note that the AI system responses may appear very fluent and well-formed but contain slight inaccuracies that are not easy to discern at first glance. Pay close attention to the text. Read it carefully as you would when proofreading.

Individual Citation:

Given a sentence and its cited passage, judge whether the passage supports the sentence. It may be helpful to ask yourself whether it is accurate to say "according to the cited passage" with the sentence following this phrase. For example, if the sentence is "Barack Obama was born in Hawaii", then ask yourself whether it is accurate to say: "According to the cited passage, Barack Obama was born in Hawaii." Be sure to check all of the information in the sentence.

You should judge the citation on a three-point scale:

- "Full Support": All of the information in the sentence is supported in the cited passage.
- "Partial Support": Only some of the information in the sentence is supported in the cited passage, but other parts of the information are missing from the cited passage.
- "No Support": The cited passage does not support any part of the statement made by the sentence.

Please provide your best judgment based on the information in the cited passage and the information stated in the sentence (when interpreted in the context of what has been said earlier in the system response). The degree of support that the cited passage provides for a sentence should be judged on its own merits; that is, the cited passage does (or does not) support the sentence regardless of whether other cited passages support the same sentence. (And if you see identical passages and identical sentences having identical interpretations, you should give them the same judgment!).

Support can be viewed as a kind of relevance, where you're asking about the "relevance" of the cited passage to the sentence (rather than the relevance of a document to a query). Like document/topic relevance, support is a somewhat subjective concept, and there may be gray areas where different people may disagree on some judgments; however, it is important that you make your decision consistently.

Note that when judging whether a cited passage supports a sentence, you should ignore whether or not the sentence and passage are relevant to the topic; support judgment is independent of topic relevance.

Tips

Every judgment you make is saved on the server automatically; you don't need to explicitly save anything.

It's important to make steady progress through your pool. If you're not sure how to judge a citation, you can skip it and come back later. You can create and apply filters to show all rows, just the rows that are Completed, or just the rows that are not Completed; this should help you find the rows that you skipped.