TREC iKAT 2024 Guidelines

Website: https://www.trecikat.com/

Guidelines v1.0

May 7, 2024

Table of Contents

Coordinators

- Mohammad Aliannejadi, University of Amsterdam
- Zahra Abbasiantaeb, University of Amsterdam
- Simon Lupart, University of Amsterdam
- Shubham Chatterjee, University of Edinburgh
- Jeff Dalton, University of Edinburgh
- Leif Azzopardi, University of Strathclyde

Participation

Participants must register to submit. To request a late registration, please email trec@nist.gov requesting a registration key. The dissemination form must be returned to submit runs.

Contact Us

- Twitter: @trec_ikat
- Email: trec.ikat.ai@gmail.com
- Google Groups: trec-ikat@googlegroups.com
- Slack: ikat-2024

Motivation

The primary goal of the TREC Interactive Knowledge Assistance Track (iKAT) is to advance the research on collaborative information-seeking conversational agents that generate personalized responses based on the information they learn about the user. iKAT is an extension of the TREC Conversational Assistance Track (CAsT).

In iKAT, the direction of the conversation not only depends on the previous responses but also on the personal information about the user. So given a topic, different users with different profiles will have different conversational trajectories and the final response of the system for the same information need will not necessarily be the same for them. With the rise of LLMs, this task is particularly timely, as the personalization of responses through relevant context, prompting, and mixed-initiative will be an emerging challenge.

Track Overview

In iKAT, the direction of the conversation can be changed in each turn based on:

1. The previous responses from the user,
2. The persona of the user, and
3. The information learned from the user (background, perspective, and context).

The persona of the user and their information needs form the direction of the conversation. Each topic will have multiple conversations based on different personas. This will mean different responses will need to be produced that demonstrate the personalized nature of the conversations. To this end, the persona and the information needs of the user are modeled by generating a Personal Textual Knowledge Base (PTKB) which is available to the system during the conversation.

Note: Similar to the first year, the PTKB is provided for each conversation and the participants do not have to generate or update it.

Task Overview

In Year 2, the following inputs are provided to the participants at each conversation turn:

1. Personal Text Knowledge Base (PTKB);
2. Conversation history (user utterance and system response for previous turns);
3. Current user's information need (user utterance).

We offer the following tasks:

- **PTKB Statement Ranking**: For each turn, given the PTKB, return a ranking of the relevant PTKB statements. This task is essentially a binary classification task. So, the output required is a list of statements from the PTKB.
- **Passage Ranking & Response Generation**: For each turn, retrieve and rank relevant passages from the given collection in response to a user utterance. Then use the ranked passages to return a set of responses. Each response may be simply a passage from the collection. Alternatively, it may also be an extracted or generated summary from one or more passage results. All responses must have at least one passage called “provenance” from the collection.
- **Only Response Generation**: For each turn, given the ranked list of passages, return a set of responses. As already specified above, each response may be simply a passage from the collection. Alternatively, it may also be an extracted or generated summary from one or more passage results. All responses must have at least one passage called “provenance” from the collection.

We will provide baseline passage ranking and response generation methods for each of the tasks.

Manual runs of the tasks, where the participants can also use the following inputs provided for each conversational turn:

4. Manually rewritten queries;
5. Manually selected relevant PTKB statements.

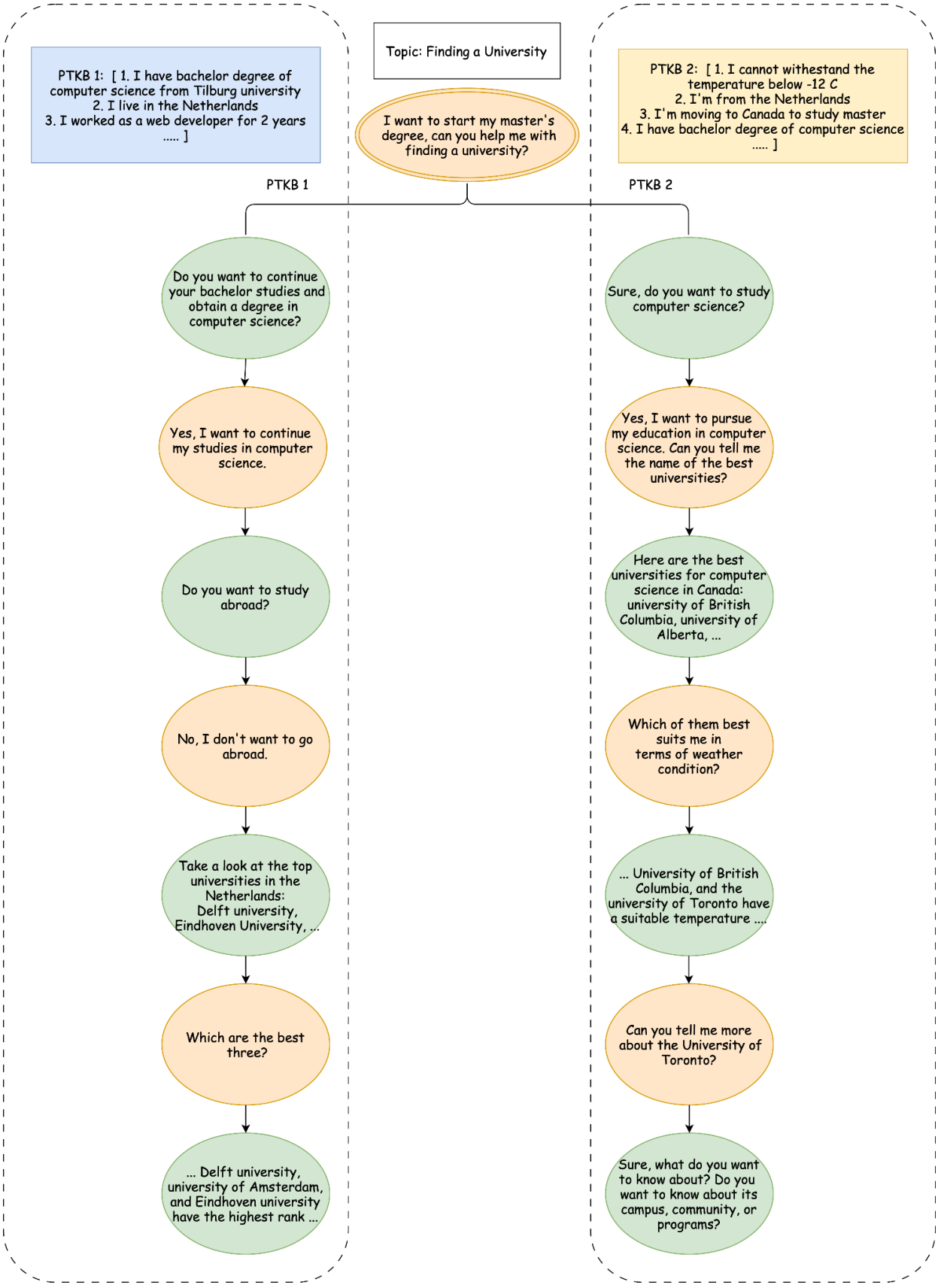
Submission Classes

- **Automatic**: These use raw utterances (with automatic rewrite/expansion) methods.
- **Manual**: These can use manual human-rewritten queries and the relevant PTKB statements.
- **Only response generation:** These use the given passage ranking for response generation. The focus is only on response generation.

Example Dialogue Tree

An example of two different conversations based on different personas for the same topic is shown in the following figure. For each user turn, systems should return a ranked list of text responses. Each response has one or more (ranked) source

passages as provenance. In addition, the systems should provide a sorted list of relevant statements of PTKB with the corresponding relevance score.


PTKB 1: [1. I have bachelor degree of computer science from Tilburg university
2. I live in the Netherlands
3. I worked as a web developer for 2 years
.....]
Topic: Finding a University
I want to start my master's degree, can you help me with finding a university?
PTKB 2: [1. I cannot withestand the temperature below -12 C
2. I'm from the Netherlands
3. I'm moving to Canada to study master
4. I have bachelor degree of computer science
.....]
PTKB 1
PTKB 2
Do you want to continue your bachelor studies and obtain a degree in computer science?
Yes, I want to continue my studies in computer science.
Do you want to study abroad?
No, I don't want to go abroad.
Take a look at the top universities in the Netherlands: Delft university, Eindhoven University, ...
Which are the best three?
... Delft university, university of Amsterdam, and Eindhoven university have the highest rank ...
Sure, do you want to study computer science?
Yes, I want to pursue my education in computer science. Can you tell me the name of the best universities?
Here are the best universities for computer science in Canada: university of British Columbia, university of Alberta, ...
Which of them best suits me in terms of weather condition?
... University of British Columbia, and the university of Toronto have a suitable temperature ....
Can you tell me more about the University of Toronto?
Sure, what do you want to know about? Do you want to know about its campus, community, or programs?

Below, we provide a detailed explanation of the above diagram.

Turn	Field	Value
Topic 1-1, PTKB_1		
1	User utterance	I want to start my master's degree, can you help me with finding a university?
	System Response	Do you want to continue your bachelor's studies and obtain a degree in computer science?
	Passage provenances	Null
	PTKB provenances	• I have a bachelor's degree in computer science from Tilburg University
2	User utterance	Yes, I want to continue my studies in computer science.
	System Response	Do you want to study abroad?
	Passage provenances	Null
	PTKB provenances	• I live in the Netherlands;
3	User utterance	No, I don't want to go abroad.
	System Response	I can help you find a university to continue your studies in the Netherlands as a computer science student. Take a look at these Top Computer Science Universities in the Netherlands: • Delft University of Technology. • Eindhoven University of Technology • Vrije Universiteit Amsterdam • University of Amsterdam • Leiden University • Radboud University • Utrecht University • University of Twente
	Passage provenances	**Clueweb22-en0034-09-03452:0**
	PTKB provenances	• I live in the Netherlands • I have a bachelor's degree in computer science from Tilburg University
4	User utterance	Which are the best three?

	System Response	Based on the ranking of research.com the top three universities are as follows: 1. Delft university of technology 2. University of Amsterdam (UVA) 3. Eindhoven University of Technology
	Passage provenances	**Clueweb22-en0016-21-09213:0**
	PTKB provenances	• I live in the Netherlands • I have a bachelor's degree in computer science from Tilburg University
Topic 1-2, PTKB_2		
1	User utterance	I want to start my master's degree, can you help me with finding a university?
	System Response	Sure, do you want to study computer science?
	Passage provenances	Null
	PTKB provenances	• I have a bachelor's degree in computer science with a GPA of 3.3 out of 4.0
2	User utterance	Yes, I want to pursue my education in computer science. Can you tell me the names of the best universities?
	System Response	Here are the best universities for computer science in Canada: 1. University of British Columbia 2. University of Alberta 3. Concordia University 4. Simon Fraser University 5. The University of Toronto
	Passage provenances	**Clueweb22-en0026-31-15538:1,** **Clueweb22-en0026-31-15538:4,** **Clueweb22-en0026-31-15538:6,** **Clueweb22-en0040-41-06056:0**
	PTKB provenances	• I have a bachelor's degree in computer science with a GPA of 3.3 out of 4.0 • I'm moving to Canada to study master
3	User utterance	Which of them best suits me in terms of weather conditions?
	System Response	University of British Columbia in Vancouver has temperatures near 80 degrees Fahrenheit (27 degrees Celsius) in summer and up to 45

		degrees Fahrenheit (about 7 degrees Celsius) in winter which is suitable for you. The University of Toronto is acceptable since it has cold winters, average temperatures can drop below -10 ° C but not below 12 degrees for long. The Concordia University in Montreal is not suitable for you since in the winter, it could reach minus 40 with the wind chill. The University of Alberta is also not suitable for you. In winter the average temperature varies between -6.5°C (20.3°F) and -13.5°C (7.7°F). Simon Fraser University is not acceptable to you. The city in which the university is located will reach temperatures of -14 in the winter.
	Passage provenances	**Clueweb22-en0000-94-02275:0,** **Clueweb22-en0027-06-08704:1,** **Clueweb22-en0005-63-12144:0,** **Clueweb22-en0013-01-17558:1,** **Clueweb22-en0014-39-04143:0**
	PTKB provenances	• I cannot withstand the temperatures below -12 for a long time, • I'm used to heavy rains in the Netherlands
4	User utterance	Can you tell me more about the University of Toronto?
	System Response	Sure, what do you want to know about? Do you want to know about its campus, community, or programs?
	Passage provenances	Null
	PTKB provenances	Null

Primary Task Details

The main aim of iKAT can be defined as **personalized retrieval-based "candidate response retrieval" in the context of the conversation and PTKB**. A detailed explanation of the subtasks is provided in the following.

PTKB Statement (provenance) Classification

- This task is considered a binary classification problem. The output is in the form of a list of the statements from PTKB that are relevant for responding to the user's utterance.
- The provenance PTKB statements can be an empty list for some responses.

Passage (provenance) Ranking

- The provided context for passage ranking per each user utterance includes:
    - A fixed set of previous utterances and responses in the preceding turns up to the current step,
    - PTKB of the user provided in the beginning of the conversation
- **Note: Using information from 1) following turns and 2) the relevant PTKB statements from previous turns are not allowed.**
- A run can have multiple responses for each turn. We will assess just the first response for each turn.
- A run must include the provenance passages for all responses in response order. The first 1000 provenances for each turn will be assessed.
- Because a response may have multiple source passages, the score of passages in the provenance list **for a response** is used to order passages in descending order.
- Each provenance that is used for generating the response should be specified.
- Each provenance is written in the format **doc_id:passage_id**.
- A FLAG [`eval_response=False`] is provided for teams that only want to focus on the ranking task without response generation.

Response generation

- Each response can be generated from multiple passages. It can be an abstractive or extractive summary of the corresponding passages.
- Each response must have one or more ranked passages as provenance from the collection used to produce it.
- A response is a text suitable for showing to the user. It should be fluent, satisfy their information needs, and not contain extraneous or redundant information.
- A response is limited to a maximum of **250 words** (as measured by the `Tokenizer` function of `spacy.tokenizer` in spaCy v3.3 library) but should vary depending on an appropriate query-response.

Response Generation: Two versions in iKAT Year 2

In the second year of iKAT, we are offering teams two distinct options for response generation:

1. **Independent Retrieval:** Teams handle the retrieval task autonomously and generate responses based on the passages they retrieve.
2. **Provided Passage Ranking:** We supply a ranked list of passages for each turn, which teams should use as the basis for generating their responses.

Collection

The text collection contains a subset of ClueWeb22 documents, prepared by the organizers in collaboration with CMU. Documents have then been split into ~116M passages. The goal is to retrieve passages from target open-domain text collections.

License for ClueWeb22-B

Getting the license to use the collection can be time-consuming and would be handled by CMU, not the iKAT organizers. Please follow these steps to get your data license ASAP:

1. Sign the license form available on the ClueWeb22 project web page: https://lemurproject.org/clueweb22/obtain.php and send the form to CMU for approval (**jlm4@andrew.cmu.edu**).
2. Once you have the license, send a mail to Andrew Ramsay <andrew.ramsay@glasgow.ac.uk> to have access to a download link with the preprocessed iKAT passage collection, and other resources (Lucene, SPLADE indexes).

Please give enough time to the CMU licensing office to accept your request.

Note.

- CMU requires a signature from the organization (i.e., the university or company), not an individual who wants to use the data. This can slow down the process at your end too. So, it's useful to start the process ASAP.
- If you already have an accepted license for ClueWeb22, you do not need a new form. Please let us know if that is the case.
- As an alternative for (2), once you have access to ClueWeb22, you can get the raw clueweb-b/ikat collection yourself with the license, and do all

passage-segmentation yourself, but we advise you to use our processed version to avoid any error.

Please do feel free to reach out to us if you have any questions or doubts about the process, so we can prevent any delays in getting the data to you.

Passage Segmentation

For assessment, we will judge provenance passages. We segment the documents in our collection into passages in a similar manner as done by the TREC Deep Learning track for segmenting MS MARCO documents into passages: First, each document is trimmed to 10k characters. Then a 10-sentence sliding window with a 5-sentence stride is used to generate the passages.

An example document with some passage segmentation is provided in TrecWeb format below for illustration purposes:

```
<DOC>
    <DOCNO>clueweb22-en0020-70-00000</DOCNO>
    <DOCHDR></DOCHDR>
    <HTML>
        <TITLE>06AECE252028B717019F2802EB065B68</TITLE>
        <URL>https://pokemon.fandom.com/wiki/Donphan</URL>
        <BODY>
            <PASSAGE id=0>
            It has thin, elongated ears held out almost perpendicular to its body. Its four short legs are each encircled with a broken, black band
                and have three claws. Behavior Donphan's behavior was mostly seen in the anime. It likes to use moves like Rollout As a form of
                attacking, becoming impossible to stop and able to demolish houses, as well as by males to establish dominance. The male Donphan have
                rolling competitions to demonstrate their toughness to the females. It has been demonstrated that Donphan has a keen sense of smell,
                capable of sniffing out a gem known as amberite. Donphan travels in herds with a ranking system. The larger its tusks, the higher its
                rank in the herd. Habitat They live in rough, mountainous terrain. Abilities Its main ability sturdy prevents it from being knocked
                out in 1 hit,while sand veil boosts its evasion in sandstorm and makes it harder to hit.
            </PASSAGE>
            <PASSAGE id=1>
            It has been demonstrated that Donphan has a keen sense of smell, capable of sniffing out a gem known as amberite. Donphan travels in herds
                with a ranking system. The larger its tusks, the higher its rank in the herd. Habitat They live in rough, mountainous terrain.
                Abilities Its main ability sturdy prevents it from being knocked out in 1 hit,while sand veil boosts its evasion in sandstorm and
                makes it harder to hit. It may also have a decent sense of smell. Gender differences Male Donphan have longer tusks then the females.
                The tusks also seem to show leadership ranks, the bigger the tusks, the higher the rank. Evolution Donphan is the evolved form of
                Phanpy, which evolves at level 25. Game info Game locations Locations Version(s) Area(s) Rarity Gold and Silver Route 28, Silver Cave,
                Victory Road(Silver Only) Uncommon Crystal Route 45 Uncommon Ruby and Sapphire Evolve Phanpy None FireRed and LeafGreen Evolve Phanpy
                None Emerald Evolve Phanpy None Diamond and Pearl Evolve Phanpy None Platinum Evolve Phanpy None HeartGold and SoulSilver Route 28, Mt
                . Silver, Victory Road(HeartGold Only) Uncommon Black and White Trade None X and Y Friend Safari Common Pokédex entries Gold It has
                sharp, hard tusks and a rugged hide.
            </PASSAGE>
        </BODY>
    </HTML>
</DOC>
```

Topic Format

We will provide several sample topics with example baseline runs for validation and testing. Below is a sample topics file with two subtrees of the same topic. Subtrees are identified by topic and subtree ID, i.e. topic 1, subtree 2 is 1-2. Also, a passage_provenance field with a list of provenance passages and a ptkb_provenance field with a list of provenance statements from PTKB, that are used for generating the response, are included.

A sample of a topic


```
[
  {
    "number": "1-1",
    "title": "Finding a University",
    "ptkb": {
      "1": "I graduated from Tilburg University.",
      "2": "I live in the Netherlands.",
      "3": "I'm allergic to peanuts.",
      "4": "I worked as a web developer for 2 years.",
      "5": "I have a bachelor's degree in computer science.",
      "6": "I like Indian food.",
      "7": "My bachelor's GPA is 5.6.",
      "8": "I'm 26 years old.",
      "9": "My TOEFL SCORE is 91.",
        "10": "My interesting bachelor courses are data structure, algorithm, data
mining, and artificial intelligence.",
      "11": "I didn't like computer architecture and logical circuits courses."
    },
    "turns": [
      {
        "turn_id": 1,
           "utterance": "I want to start my master's degree, can you help me with
finding a university?",
         "resolved_utterance": "I want to start my master's degree, can you help me
with finding a university?",
          "response": "Do you want to continue your bachelor's studies and obtain a
degree in computer science?",
        "ptkb_provenance": [
          5
        ],
        "response_provenance": [],
        "sample_passage_ranking": [
          "clueweb22-en0034-09-03452:1",
          "clueweb22-en0034-09-03452:3",
          "clueweb22-en0034-09-03452:5",
          "clueweb22-en0034-09-03452:7",
          "clueweb22-en0034-09-03452:9"
        ]
      },
      {
        "turn_id": 2,
        "utterance": "Yes, I want to continue my studies in computer science.",
           "resolved_utterance": "Yes, I want to continue my studies in computer
science.",
         "response": "Do you want to study in the Netherlands, Europe, or somewhere
further away?",
        "ptkb_provenance": [
          2
        ],
        "response_provenance": [],
        "sample_passage_ranking": [
          "clueweb22-en0034-09-03452:2",
          "clueweb22-en0034-09-03452:4",
          "clueweb22-en0034-09-03452:6",
          "clueweb22-en0034-09-03452:8",
          "clueweb22-en0034-09-03452:10"
        ]
      },
```

```
      {
        "turn_id": 3,
        "utterance": "I'd like to stay here.",
        "resolved_utterance": "I'd like to stay in the Netherlands.",
          "response": "I can help you with finding a university for continuing your studies in the Netherlands as a computer science student. Take a look at these Top Computer Science Universities in the Netherlands: Delft University of Technology, Eindhoven University of Technology, Vrije Universiteit Amsterdam, University of Amsterdam, Leiden University, Radboud University, Utrecht University, University of Twente",
        "ptkb_provenance": [
          5,
          2
        ],
        "response_provenance": [
          "clueweb22-en0034-09-03452:1"
        ],
        "sample_passage_ranking": [
          "clueweb22-en0012-00-00012:0",
          "clueweb22-en0012-00-00012:1",
          "clueweb22-en0012-00-00012:2",
          "clueweb22-en0012-00-00012:3",
          "clueweb22-en0012-00-00012:4"
        ]
      }
    ]
  },
  {
    "number": "1-2",
    "title": "Finding a university",
    "ptkb": {
      "1": "I don't like crazy cold weather.",
      "2": "I don't have a driver's license.",
      "3": "I plan to move to Canada.",
      "4": "I'm from the Netherlands.",
      "5": "I'm used to heavy rains in the Netherlands.",
      "6": "I graduated from UvA.",
      "7": "I have bachelor's degree in computer science.",
      "8": "I speak English fluently."
    },
    "turns": [
      {
        "turn_id": 1,
           "utterance": "I want to start my master's degree, can you help me with finding a university?",
         "resolved_utterance": "I want to start my master's degree, can you help me with finding a university in Canada?",
        "response": "Sure, do you want to study computer science?",
        "ptkb_provenance": [
          7,
          3
        ],
        "response_provenance": [],
        "sample_passage_ranking": [
          "clueweb22-en0040-41-06056:0",
          "clueweb22-en0040-41-06056:1",
          "clueweb22-en0040-41-06056:2",
          "clueweb22-en0040-41-06056:3",
          "clueweb22-en0040-41-06056:4"
        ]
      },
      {
```

```
        "turn_id": 2,
         "utterance": "Yes, I want to pursue the same major. Can you tell me the name of the best universities?",
          "resolved_utterance": "Yes, I want to pursue computer science. Can you tell me the name of the best computer science universities in Canada?",
          "response": "Here are the top universities for computer science in Canada: 1) University of British Columbia, 2) University of Alberta, 3) Concordia University, 4) Simon Fraser University, 5) The University of Toronto",
        "ptkb_provenance": [],
        "response_provenance": [
          "clueweb22-en0026-31-15538:1",
          "clueweb22-en0026-31-15538:4",
          "clueweb22-en0026-31-15538:6",
          "clueweb22-en0040-41-06056:0"
        ],
        "sample_passage_ranking": [
          "clueweb22-en0010-22-22210:0",
          "clueweb22-en0010-22-22210:1",
          "clueweb22-en0010-22-22210:2",
          "clueweb22-en0010-22-22210:3",
          "clueweb22-en0010-22-22210:4"
        ]
      },
      {
        "turn_id": 3,
        "utterance": "Which of them best suits me in terms of weather conditions?",
         "resolved_utterance": "Which of the following universities best suits me in terms of weather conditions? 1) the University of British Columbia, 2) the University of Alberta, 3) Concordia University, 4) Simon Fraser University, and 5) The University of Toronto.",
          "response": "I know you don't like very cold weather, but can you give me an estimation of the temperature that is acceptable for you?",
        "ptkb_provenance": [
          1,
          5
        ],
        "response_provenance": [],
        "sample_passage_ranking": [
          "clueweb22-en0030-30-30030:0",
          "clueweb22-en0030-30-30030:1",
          "clueweb22-en0030-30-30030:2",
          "clueweb22-en0030-30-30030:3",
          "clueweb22-en0030-30-30030:4"
        ]
      }
    ]
  }
]
```


Note that the topic file contains the full conversational context for each turn requiring a system response.

Sample Topics

We are releasing two sample conversations on the topic "Finding a University". Each conversation pertains to a different persona, and the conversation is personalized to that persona. You can find the sample topic here.

Those eager to start developing and experimenting can also use the data from previous years of TREC CAsT and TREC iKAT 2023.

Note. TREC CAsT did not include a PTKB but you can be creative and modify the data according to your needs. Also, TREC CAsT used different collections (Wikipedia, KILT, MS MARCO, etc.) at different stages. ***iKAT uses a subset of the recently released ClueWeb22-B collection***.

Task Submissions

Participants submit the output of their system on the specified “test” topics. A single participant can submit maximum of:

- 4 automatic runs,
- 2 manual runs,
- 2 only response generation runs.

In the automatic runs, the participants can include response generation based on their or ranking. But **this is not mandatory**.

In the **only response generation** run, the participants **must use** the given passage provenances.

We have three submission classes for each of the 1) automatic, 2) manual, and 3) only response generation tasks. An example of the submission template for each run is below.

A sample of submission for the main task

```
{
  "run_name": "sample_run",
  "run_type": "automatic",
  "eval_response": True,
  "turns": [
    {
```

```
      "turn_id": "1-2_3",
      "responses": [
        {
          "rank": 1,
          "text": "The University of British columbia in Vancouver has temperatures near 80 degrees Fahrenheit (27 degrees Celsius) in summer and up to 45 degrees Fahrenheit (about 7 degrees Celsius) in winter which is suitable for you. The university of Toronto is acceptable since has cold winters, average temperatures can drop below -10 ° C but not below 12 degrees for long. The Concordia university in Montreal is not suitable for you since in the winter, could reach minus 40 with the wind chill. University of Alberta is also not suitable for you. In winter the average temperature varies between -6.5°C (20.3°F) and -13.5°C (7.7°F). Simon Fraser university is not acceptable for you. The city which the university is located in will reach temperatures of -14 in the winter.",
          "ptkb_provenance": [1,2],
          "passage_provenance": [
            {
              "id": "clueweb22-en0000-94-02275:0",
              "score": 0.6,
              "used": False

            },
            {
              "id": "clueweb22-en0027-06-08704:1",
              "score": 0.5,
              "used": True
            },
            {
              "id": "clueweb22-en0005-63-12144:0",
              "score": 0.4,
              "used": False
            },
            {
              "id": "clueweb22-en0013-01-17558:1",
              "score": 0.38,
              "used": True

            },
            {
              "id": "clueweb22-en0014-39-04143:0",
              "score": 0.3,
              "used": False
            }
          ]
        }
      ]
    }
  ]
}
```


A sample of submission for the only response generation task

```
{
  "run_name": "sample_run",
  "run_type": "only_response",
  "turns": [
    {
      "turn_id": "1-2_3",
      "responses": [
        {
          "rank": 1,
         "text": "The University of British columbia in Vancouver has temperatures near 80 degrees Fahrenheit (27 degrees Celsius) in summer and up to 45 degrees Fahrenheit (about 7 degrees Celsius) in winter which is suitable for you. The university of Toronto is acceptable since has cold winters, average temperatures can drop below -10 ° C but not below 12 degrees for long. The Concordia university in Montreal is not suitable for you since in the winter, could reach minus 40 with the wind chill. University of Alberta is also not suitable for you. In winter the average temperature varies between -6.5°C (20.3°F) and -13.5°C (7.7°F). Simon Fraser university is not acceptable for you. The city which the university is located in will reach temperatures of -14 in the winter.",
          "ptkb_provenance": [1,2],
          "passage_provenance": [
            {
              "id": "clueweb22-en0000-94-02275:0",
              "used": True
            },
            {
              "id": "clueweb22-en0027-06-08704:1",
              "used": True
            },
            {
              "id": "clueweb22-en0005-63-12144:0",
              "used": False
            },
            {
              "id": "clueweb22-en0013-01-17558:1",
              "used": False
            },
            {
              "id": "clueweb22-en0014-39-04143:0",
              "used": True
            }
          ]
        }
      ]
    }
  ]
}
```


A sample of submission for the manual task


```
{
  "run_name": "sample_run",
  "run_type": "manual",
  "eval_response": True,
```

```
  "turns": [
    {
      "turn_id": "1-2_3",
      "responses": [
        {
          "rank": 1,
         "text": "The University of British columbia in Vancouver has temperatures near 80 degrees Fahrenheit (27 degrees Celsius) in summer and up to 45 degrees Fahrenheit (about 7 degrees Celsius) in winter which is suitable for you. The university of Toronto is acceptable since has cold winters, average temperatures can drop below -10 ° C but not below 12 degrees for long. The Concordia university in Montreal is not suitable for you since in the winter, could reach minus 40 with the wind chill. University of Alberta is also not suitable for you. In winter the average temperature varies between -6.5°C (20.3°F) and -13.5°C (7.7°F). Simon Fraser university is not acceptable for you. The city which the university is located in will reach temperatures of -14 in the winter.",
          "passage_provenance": [
            {
              "id": "clueweb22-en0000-94-02275:0",
              "score": 0.6,
              "used": False

            },
            {
              "id": "clueweb22-en0027-06-08704:1",
              "score": 0.5,
              "used": True
            },
            {
              "id": "clueweb22-en0005-63-12144:0",
              "score": 0.4,
              "used": False
            },
            {
              "id": "clueweb22-en0013-01-17558:1",
              "score": 0.38,
              "used": True

            },
            {
              "id": "clueweb22-en0014-39-04143:0",
              "score": 0.3,
              "used": False
            }
          ]
        }
      ]
    }
  ]
}
```


1. The **run_name** is a run submission identifier that should be descriptive and unique to your team and institution.
2. The **run_type** is one of the two types listed above, *automatic* and *manual*.

3. Each **turn** in the **turns** list should contain a **turn_identifier**, consisting of the **topic_id-subtree_id** and **turn_id concatenated** with an **underscore**, e.g. 1-2_3 for topic 1, subtree 2, and turn 3.
4. Each **turn** should also contain a list of **responses**. A **response** consists of **text** and a **provenance** list. Each **provenance** should have an **ID**, **text**, **used**, and **score**. The **used** field indicates whether the passage is used for response generation or not.
5. Each **turn** includes a **sorted list of statements from PTKB** based on the **relevance score** of **each statement** from PTKB to the current turn.
6. If you want to only do the retrieval and don't want to do the response generation in an automatic run, you can leave the **'text'** field empty and change the **'eval_response'** to False.

For provenance ranking, this will be converted to a traditional TREC run format:

```
31_1-1 Q0 clueweb22-en0000-94-02275:0 1 0.5 sample_run
```

Runs may include up to 1000 responses for each user turn. For provenance ranking, only the first 1000 pieces of unique provenance will be used. As in previous year of iKAT, only limited top-*k* responses and provenances will be assessed according to resource constraints.

Evaluation

We will use the relevance assessment methods used in the previous year of iKAT 2023 for relevance to individual turns.

1. **Provenance Passage Assessment**: The provenance passages that are used to produce the responses will be pooled and assessed. The relevance scale will be the same as in previous years of CAsT and the first year of iKAT, see the previous overview papers for details. The standard ranking metrics such as P@k, NDCG@k, and MAP will be calculated using the judgments. We will focus on the earlier positions (1, 3, 5).
2. **Response Assessment:** A response may be a simple passage or a summary of one or more passages. We will assess the top-ranked response (or top-k) from all systems for a subset of turns. Only responses with at least one relevant

provenance passage will be judged. Responses will be assessed for relevance, naturalness, completeness, and groundedness. The responses and judgments on them will be released with the judgments.

3. **Extracted Relevant Statements Assessment:** The standard metrics like Precision, Recall, P@k, and MAP will be used for evaluating the sorted list of relevant statements from PTKB.

Similar to iKAT Year 1, only a subset of turns may be evaluated for provenance ranking effectiveness. This will be disclosed to participants after the assessment is completed.

Timeline

Task	**Date**
Guidelines released	May 2024
Test topics released	June 2024
Submission deadline	31 August 2024
Results released to participants	October 2024