

ARQMath@CLEF 2020

Answer Retrieval for Questions on Math

Guidelines

Version 3, May 11th, 2020 (R. Zanibbi)

Task coordinators: Behrooz Mansouri, Anurag Agarwal, Douglas W. Oard, and Richard Zanibbi.

Changes (from Version 2)

- Added [note on valid hits](#) for Tasks 1 and 2

Overview

This document provides guidelines for participation in the ARQMath 2020 shared task at CLEF 2020, which is the first time that the task will be run. The **schedule and deadlines** for the task may be found on the [ARQMath web page](#), including any updates.

Registration. Registration for the task should be done online through CLEF 2020, at <http://clef2020-labs-registration.dei.unipd.it>. Registration closes on April 26, 2020, which we've been told is a firm deadline.

Communication. Communication between organizers and participants is being arranged through Google Groups, in the [ARQMath forum](#). Please visit the forum link to request access and join in discussions about the task. Additional information about the task is available on the [ARQMath web page](#).

Tasks and Evaluation Protocols. There are two tasks. In the first task, a question post is provided as a query, and participant systems search a collection of question and answer posts for answers to the question. In the second task, a formula within a post acts as the query, and formulas in posts from the collection are returned.

The relevance of retrieved formulas will be considered in the context of their associated post. Details of the **evaluation protocols for tasks 1 and 2** may be found in the evaluation protocols document in the [ARQMath Google Drive](#).

- **Valid Hits:** Note carefully the following constraints for Tasks 1 and 2. **Hits that do not adhere to these constraints will be filtered during assessment.**
 - **Task 1:** The *Post_id* field for each hit must be an **answer post** (i.e., it cannot be a question post, a comment, or any other element type).

- **Task 2:** The *Post_id* field for each *Formula_Id* in the results must belong to a **question post or answer post** (and not a comment or other element type). The provided formula index files contain the *Post_id* and post type associated with each formula in the collection.

Collection and Topics. The collection and topics have been posted on the [ARQMath Google Drive directory](#). The collection is adapted from the community question-answering forum [Math Stack Exchange](#), whose data is provided free for non-commercial use as snapshots on the Internet Archive. The snapshot used is from mid-2019. The main collection consists of question and answer posts from 2010-2018. **Additional details about the collection can be found in the README files provided with the collection.**

The question posts used for topics are taken from 2019. The formula queries for Task 2 are selected from question posts used for topics in Task 1.

Runs. Manual and automatic runs are permitted for both tasks. In a manual run, results may be produced by any means, including means that involve manual intervention for query formulation, result selection, or any other purpose. Automatic runs must be produced algorithmically by computer, without user intervention.

Participants should provide run in tab-separated variable (TSV) format, and name their run files using the following convention:

```
[group]-[task]-[run-type]-[data-used]-[eval].tsv
* [group]: CLEF-registered team name.
* [task]: task1 / task2 / task2NOC (see note below)
* [run-type]: manual / auto
                (manual or automatic)
* [data-used]: text / math / both
                (text-only, formula-only, formula and text)
* [eval]: P / A
                (primary run, alternate run)
```

```
Examples: TeamX-task1-auto-both-P.tsv
          TeamX-task1-auto-text-A.tsv
          TeamA-task2-manual-both-P.tsv
          TeamA-task2NOC-auto-math-A.tsv
```

Run file contents should be formatted as described in the [ARQMath2020-EvaluationProtocols](#) document.

Submitting Runs. Runs should be uploaded to the Google Drive directory that will be provided for participants, using the appropriate naming of run files as given above.

Up to a **maximum of 5 runs** may be submitted for evaluation by each group for Tasks 1 and 2. **We guarantee that the run identified as 'primary' will contribute to the judgment pools**, with the inclusion of alternate runs

in the judgment pools depending upon the number of submissions. **Participants should therefore make sure to submit a *single* 'best' run as primary (see file naming convention above), to insure that they are included in the evaluation.**

IMPORTANT Note Regarding Task 2. Participants should index formulas using the provided formula identifiers in the ARQMath corpus. These will be used to evaluate formulas in-context (i.e., within their associated post and thread). **If you *do not* index formulas in their original context for Task 2, indicate this using the task label 'task2NOC' in the associated run file name (see run file naming requirements above). Note that these runs must provide valid ARQMath formula and post identifiers to be evaluated for Task 2.**

