File Structure

Files

1. **qrel\_task2\_2022\_official\_evaluation.tsv**: This file is the qrel that should be used for official evaluation of ARQMath-2 dataset and comparison with previous works. There are 76 topics in this file.
2. **qrel\_task2\_2022\_additional.tsv:** This file contains 14 additional topics.
3. **qrel\_task2\_2022\_all.tsv:** In the qrel\_all file, all the test and train topics are put together in a single file, for purposes such as using all topics for cross-validation.

Directory

1. **Qrel Files With Formula Ids (Not for official Evaluation)**

This directory contains 3 files:

1. **qrel\_task2\_2022\_formula\_id\_official.tsv:** This file shows the relevance score given to each pair of topic and formula id. It contains only the topics used in the test set. These are the topics used for the official ARQMath-3 evaluation.
2. **qrel\_task2\_2022\_formula\_id\_additional.tsv:** This file shows the relevance score given to each pair of topic and formula id. Additional topics can be used only for training purposes.
3. **qrel\_task2\_2022\_formula\_id\_all.tsv:** combination of the two previous files.

FQA:

1. **Which file to use?**

To evaluate the retrieval results, the only file that should be used is “**qrel\_task2\_2022official\_evaluation.tsv**”. This file has pairs of topic and visual\_id (not formula\_id). Note that the official evaluation of ARQMath-3 is done based on visual ids.

For training purposes, one might use “**qrel\_task2\_2022\_additinoal.tsv**” or “**qrel\_task2\_2022\_all.tsv**”.

1. **How to use QREL files with Formula Ids?**

These files can be used for further analysis on the actual assessed topics. They contain the assessed hits after the pooling was done.

1. **What is the relationship between qrel files with formula ids and visual ids?**

The official qrel files (with visual ids) were created based on qrel files with formula ids. The qrel files with formula id shows what was actually assessed in ARQMath; showing the formula id that were in the pool.

After the assessment was done, each formula id was mapped to its visual id. If there were multiple instances of a formula, the **maximum** relevance score given to instances is considered. One might use a different function instead of max.