ARQMath-3 official assessment was done on the test set of 78 topics. The results should be reported only on the “**qrel\_task3\_2022\_official.tsv**” file.

However, there are 9 additional topics assessed. 2 of these topics are those that were assessed before creating the test qrel, and were removed due to having a lower number of relevant hits.

In the “qrel\_task3\_2022\_all.tsv” file, all the test and train topics are put together in a single file.

The file “task3-extra-assessment.tsv” provides the assessment done regarding two aspect of generated answers:

1- If they are machine generated or not?

2- If they have non-relevant information in them?