**Version V1.0**

The version V1.0 was the official version used for ARQMath 2020.

**Version V2.0**

The version V2.0 is a more complete version where more SLT and OPT representations are available.

The completion of this version has been done using the code and conversions made available by “Vítek Novotný” in <https://github.com/MIR-MU/ARQMath-data-preprocessing.git>.

The newer version of LatexML 0.8.5, release in 2020-11-17. The error rate in conversion of SLTs reduced from 8% to 0.13%. Also, a new attribute is added to index files, that shows the visual ids of each formula.

**Version V3.0**

The changes in V3.0 includes:

1- As mentioned in CLEF 2021, some of the formulas had encoding issues. For example formula a>b was represented as a&gt;b. This caused issue with SLT and OPT representation as &gt; was considered as a text. This issue is addressed in V3.0.

2- For some formulas, the extracted SLT and OPT representations for some formulas were wrong due to wrong handling of results from LaTeXML. This issue is addressed in V3.0.

3- Some formulas had wrong visual id assigned to them. This issue was fixed in V3.0. For this we have three columns. **old\_visual\_id** shows the visual id used in ARQMath-2. **new\_visual\_id** shows the visual id is the corrected visual\_id (and if there was no issue, the old visual id) and finally, **issue** that value ‘v’ shows the visual id has changed.

4- There were formulas in the TSV index file that were not present in XML files. We fix this issue by adding a column **issue,** using value ‘d’ to indicate that the formula is not in the XML file. Note that if formula does not exist in XML files and the visual id has changed, we will use ‘dv’ in the **issue** column.

5- Formulas in comments did not have the comment id. We included a column named **comment\_id** that has the comment id in which the formula appeared.