

Code pieces for the reproducibility of experiments for the submission “Quality Estimation for Corpus Filtering”. We share all the documents with an anonymous google account to preserve anonymity. The notebooks are for the zh-en changing the language pair names should allow testing different language pairs as well. We will also try to provide a pro account for this email address to allow for the usage of GPU's when testing. But this might not be possible while preserving anonymity.

MultiQE - Multitask Training

https://colab.research.google.com/drive/1Yx7sVBI130lyNNdGs__JlvJNPKF_ptFI?usp=sharing

MultiQE - Translation

<https://colab.research.google.com/drive/1G2DQuzE6GD95WlWmC9d1rOFD7ScaIDR0?usp=sharing>

MultiQE - Embedding

https://colab.research.google.com/drive/1v_pvGbkg02FBgXDS78UM-m3uyAnvJPYk?usp=sharing

MultiQE - TATOEBA

<https://colab.research.google.com/drive/19eABs05mUnq8lL-Skeptx2Xi04xY-Pbz?usp=sharing>

MultiQE - BUCC

https://colab.research.google.com/drive/1dqIaQfko2cFmHlvhBlczOPsXNTwk6Jz_?usp=sharing