**Automatic and Manual Metrics for Operational Translation Evaluation**

[**http://mte2014.github.io/**](http://mte2014.github.io/)

**26 May 2014**

Workshop at Language Resources and Evaluation Conference (LREC) 2014

<http://lrec2014.lrec-conf.org>

**In brief:**

We invite you to join us for an interesting day of work (and play!) as we discuss metrics for machine translation quality assessment and participate in some hands-on task-based translation evaluation.

This workshop on Automatic and Manual Metrics for Operational Translation Evaluation (MTE 2014) will be a full-day LREC workshop to be held on Monday, May 26, 2014 in Reykjavik, Iceland. The format of MTE 2014 will be interactive and energizing: a half-day of short presentations and discussion of recent work on machine translation quality assessment, followed by a half-day of hands-on collaborative work with MT metrics that show promise for the prediction of task suitability of MT output. The afternoon hands-on work will follow from the morning’s presentations, with some of the hands-on exercises developed directly from the submissions to the workshop.

**Details:**

While a significant body of work has been done by the machine translation (MT) research community towards the development and meta-evaluation of automatic metrics to assess overall MT quality, less attention has been dedicated to more operational evaluation metrics aimed at testing whether translations are adequate within a specific context: purpose, end-user, task, etc., and why the MT system fails in some cases. Both of these can benefit from some form of manual analysis. Most work in this area is limited to productivity tests (e.g. contrasting time for human translation and MT post-editing). A few initiatives consider more detailed metrics for the problem, which can also be used to understand and diagnose errors in MT systems. These include the Multidimensional Quality Metrics (MQM) recently proposed by the EU F7 project QTLaunchPad, the TAUS Dynamic Quality Framework, and past projects such as FEMTI, EAGLES and ISLE. Some of these metrics are also applicable to human translation evaluation. A number of task-based metrics have also been proposed for applications such as topic ID / triage and reading comprehension. The purpose of this workshop is to bring together representatives from academia, industry and government institutions to discuss and assess metrics for manual and automatic quality evaluation, with an eye toward how they might be leveraged or further developed into task-based metrics for more objective “fitness for purpose” assessment. We will also consider comparisons to well-established metrics for automatic evaluation such as BLEU, METEOR and others, including reference-less metrics for quality prediction. The workshop will benefit from datasets already collected and manually annotated for translation errors by the QTLaunchPad project (<http://www.qt21.eu/launchpad/>) and will cover concepts from many the metrics proposed by participants in the half-day of hands-on tasks.

**Up-to-the-minute information and (most importantly) Registration:**

Additional details and schedule will be posted at the workshop website<http://mte2014.github.io/> as they become available. Register to attend via the LREC registration site at<http://lrec2014.lrec-conf.org/en/registration/>.

We look forward to seeing you there!

The MTE 2014 Organizing Committee

*Keith J. Miller (MITRE)*

*Lucia Specia (University of Sheffield)*

*Kim Harris (GALA and text & form)*

*Stacey Bailey (MITRE)*