



Okapi+QuEst: Translation Quality Estimation within Okapi

Gustavo Henrique Paetzold, University of Sheffield, ghpaetzold1@sheffield.ac.uk

Lucia Specia, University of Sheffield, l.specia@sheffield.ac.uk

Yves Savourel, ENLASO, ysavourel@enlaso.com

<https://bitbucket.org/okapiframework/quest>

Description

Due to the ever growing applicability of machine translation, estimating the quality of translations automatically has become a necessary task in various scenarios, for example, when deciding whether a machine translation is good enough for human post-editing. This demonstration presents the outcome of a collaborative project between the University of Sheffield and ENLASO, funded by EAMT, the European Association for Machine Translation. The project aimed to integrate a lightweight and user-friendly version of QuEst (<http://www.quest.dcs.shef.ac.uk/>) – a quality estimation toolkit, into Okapi (<http://www.opentag.com/okapi/>) – a framework with various components and applications designed to help create and improve translation and localisation processes. As result, Okapi users are now offered a software plugin to build and apply quality estimation models for translations produced within the framework. In addition to the standard functionalities of QuEst, the project involved the creation of new methods to facilitate the generation of the linguistic resources necessary for building quality estimation models.

When installed in the Okapi Framework, the QuEst plugin introduces three “steps” to the Okapi Pipeline, which includes 75 other steps (for example, a Moses translation step). Users can create translation quality estimation tasks by adding the relevant QuEst steps to their pipeline. The three steps provided by the QuEst plugin in Okapi are:

- **SVM Model Builder step:** Provides an easy way for users to extract features from texts and train translation quality estimation models using the LibSVM tool. These models can then be used by the Quality Estimation step.
- **Quality Estimation step:** Allows users to apply an existing quality estimation model or a model created by the SVM Model Builder step to produce quality estimation scores for new translations.
- **Properties Setting step:** Gathers the quality estimation scores produced by the Quality Estimation step and place them into an XLIFF which can be interpreted by annotation tools such as Ocelot (<http://open.vistatec.com/>) and used in annotation tasks, such as quality inspection by humans.

The Okapi version with QuEst can be downloaded from <https://code.google.com/p/okapi-quest/>. The tool is free, open-source and cross-platform (Java). It is easy to install and is provided with clear documentation through wikipages with step-by-step tutorials. Different from the standalone version of QuEst, it produces most necessary linguistic resources automatically to help inexperienced users. Following Okapi's tradition, the tool offers a graphical interface, making it easier to use.