The ACCEPT Academic Portal: A User-Centred Online Platform for Pre-editing and Post-editing

GULATI, Asheesh, et al.

Abstract

The ACCEPT Academic Portal is a showcase of the ACCEPT project, an FP7 project aimed at improving statistical machine translation of community content through minimally-intrusive pre-editing techniques, statistical machine translation improvement methods and post-editing strategies. The project’s technology was originally developed as a series of plug-ins for integration into different interfaces such as forums, web portals or crowdsourcing platforms by means of APIs. To make the ACCEPT technology accessible to a wider public and, in particular, to teachers and students, we undertook the task of transforming the existing demo portal into an easy-to-use, fully-integrated online platform that combines the typical modules of an MT workflow and is specially designed for academic purposes.

Reference


Available at:
http://archive-ouverte.unige.ch/unige:73249

Disclaimer: layout of this document may differ from the published version.
The ACCEPT Academic Portal is a showcase of the ACCEPT project, an FP7 project aimed at improving statistical machine translation of community content through minimally-intrusive pre-editing techniques, statistical machine translation improvement methods and post-editing strategies. The project’s technology was originally developed as a series of plug-ins for integration into different interfaces such as forums, web portals or crowdsourcing platforms by means of APIs. To make the ACCEPT technology accessible to a wider public and, in particular, to teachers and students, we undertook the task of transforming the existing demo portal into an easy-to-use, fully-integrated online platform that combines the typical modules of an MT workflow and is specially designed for academic purposes.

**THE SETTINGS PAGE**
- Available language pairs: French <-> English, English > German.
- Use of sample data (community content), custom data or data inserted in the inline editor.
- Integrated modules: Pre-editing, Machine Translation, Post-editing and Evaluation. Some modules may be skipped.

**THE PRE-EDITING MODULE**
- Checking module to verify compliance of input data with pre-editing rules:
  - Interactive or automatic rule application;
- Possibility of ignoring rules or adding words to the system to avoid them in subsequent checks.

**THE MT MODULE**
- Machine translation of the raw and pre-edited version of the input data.
- Comparison between output translations (mark up of differences).
- Phrase-based Moses system trained on TMs supplied by Symantec + Europarl and news-commentary data.
  
  Potential developments: other SMT and RB systems.

**THE POST-EDITING MODULE**
- Sentence-level post-editing interface, visualisation of the source, the MT output, and the sentence currently being edited.
- Post-editing with or without post-editing rules (human or machine-oriented).
- Mark up of changes produced by the rules.

**SUMMARY AND STATISTICS**
- Summary and statistics of each step of the process:
  - *Pre-editing*: applied rules, ignored rules, changed tokens;
  - *MT*: system used;
  - *Post-editing*: time, keystrokes, changed tokens.

Try it on [http://isslnx1.unige.ch/aap](http://isslnx1.unige.ch/aap)!