

ACCEPT

Automated Community Content Editing PorTal (ACCEPT)

Contact: Pierrette Bouillon

Pierrette.bouillon@unige.ch

<http://www.accept.unige.ch>

ACCEPT is a Collaborative Project – STREP aimed at developing new methods and techniques to make machine translation (MT) work better in the environment characterised by internet communities sharing specific information. Today, anyone can in principle create information and make it available to anyone in the world with Internet access. Yet the language barrier remains: however accessible information is, it is still only available to those who speak the language it is written in. ACCEPT's mission is to help communities share information more effectively across the language barrier by improving the quality of machine-translated community content. The approach consists of the following main axes of research and development:

- Development of user-friendly (minimally intrusive) strategies for pre-editing the content for statistical machine translation. The project will identify the most important types of corrections that need to be applied to the source content in order to attain a higher translation quality.
- Development of strategies for post-editing. Ideally, post-editing of the translation results is done by bilingual skilled experts, but the lack of such experts is a major bottleneck. To overcome this bottleneck, the project will develop post-editing strategies which do not require proficiency of the source language, but only of the target language, thus enlarging the pool of (volunteer) skilled experts.
- Improvement of learning and development of feedback loops to improve Statistical Machine Translation (SMT) for community data. SMT systems can learn well from large amounts of similar content in a single domain, but improvement is necessary in areas where resources (parallel data) are sparse and heterogeneous. The project will develop innovative domain adaptation methods and will use linguistic information to cope with these issues. Moreover, it will take into account feedback from the post-editing process to automate corrections whenever possible. Another novel research topic that will be addressed is the use of text analytics for SMT. The project will try to determine if what we know about the content can help produce better translations (for instance, translations that preserve sentiment polarity).

For the first time, pre-editing, MT and post-editing will be linked together not just in a process, or workflow, but by connecting the software components together and by developing new linguistic software components specifically optimised for community content translation.

The ACCEPT project will be addressing the challenge of removing the language barrier in two slightly different scenarios: firstly, for content in a typical commercial product forum relating to Symantec network security products (<http://community.norton.com/norton/>), and secondly, for content in the community of volunteer translators of TWB – Translators Without Borders, an NGO that creates medical, educational and nutritional information for use by people in areas of need (<http://translatorswithoutborders.org/>). Enabling effective MT for the information which is often created by subject-matter experts rather than professional writers will significantly increase the reach of this information, and, in particular, help Translators Without Borders better achieve its mission of saving lives by delivering critical information in the right language at the right time.