



Challenge

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.

Existing machine translation (MT) engines cannot produce acceptable results for community content, due to its poor linguistic quality (deviation from spelling, punctuation and grammar norms; use of jargon, abbreviations, informal tone; and lack of language proficiency).

ACCEPT's mission is to help communities share information more effectively across the language barrier, by improving the quality of machine-translated community content.

Facts

Duration: 36 months
Start date: 01/01/2012
Finish date: 31/12/2014
EU funding: €1,825,000
EU project officer:
Pierre-Paul Sondag
Website:
www.accept.unige.ch

Innovation

The ACCEPT project proposes a new approach to help MT work better for community content, in order to ensure that the result is comprehensible and correct. 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, allowing to identify the most important types of corrections that lead to a higher translation quality
- Development of strategies for post-editing, by capitalizing on subject matter experts proficient in the target language rather than on bilingual experts, who are rare and expensive
- Improvement of learning and development of feedback loops to improve statistical machine translation (SMT) for community data: innovative domain adaptation and linguistic back-off methods will be used to deal with the sparseness and heterogeneity of parallel data; feedback from the post-editing process as well as mechanisms for text analytics will be taken into account to improve the statistical MT engines.

For the first time, pre-editing, MT and post-editing will be linked together in a process and workflow specifically optimised for community content translation.

Impact

The technology built will help drive take-up of MT in the Web 2.0 paradigm. The ACCEPT project will focus on two slightly different scenarios, a typical commercial product forum and the community of volunteer translators Translators Without Borders (TWB).

ACCEPT will allow citizens across the EU better access to communities in both commercial and non-profit environments. It will also make companies in Europe better able to engage with their customers across the language barrier, making them more competitive and allowing them to expand more quickly across Europe. This will bring appreciable industrial and commercial impact. The social impact of the project is highlighted in the participation of TWB, whose main focus is helping European NGOs provide better services in multilingual theatres of operations.

User Group

- Symantec product forums
- Translators Without Borders

Consortium

- Université de Genève, Switzerland
- University of Edinburgh, United Kingdom
- Acrolinx GmbH, Germany
- Symantec Limited, Ireland
- Lexcelera, France

Coordinator contact:

Pierrette Bouillon • Tel. : +41 22 379 86 79 • Fax: +41 22 379 86 89 • E-mail: Pierrette.Bouillon@unige.ch