TaaS – Terminology as a Service

EU Seventh Framework Programme (FP7)
ICT Call - SME initiative on Digital Content and Languages
Small or medium scale focused research project (STREP)
Project ID: 296312
http://www.taas-project.eu

List of partners
TILDE Tilde, Latvia (coordinator)
Cologne University of Applied Sciences, Germany
KILGRAY Kilgray, Hungary
The University Of Sheffield. University of Sheffield, United Kingdom
TAUS TAUS, Netherlands

Project duration: May 2012 — May 2014

Summary

The project implements a new paradigm in terminology work creating an online platform termunity.com to automate terminology identification, acquisition and processing tasks. The automation of individual tasks is provided as a set of interoperable cloud-based services integrated into workflows. These services automate identification of term candidates in user-provided documents, the lookup of translation equivalents in online terminology resources and on the Web by automatically extracting multilingual terminology from comparable and parallel online resources. Although term identification is very challenging even to human annotators, we can achieve a comparable precision with automatic methods using the extended term tagging system. For example, for Latvian an average precision of 53.8% was achieved in comparison to an average annotator agreement rate of 63.3%. Collaborative involvement of users contributes to refinement and enrichment of raw terminological data. An API is provided for usage of the terminology services and terminology data by external systems. This API-level integration is currently implemented by the memoQ CAT tool and the LetsMT statistical MT system. In the framework of the project several methods have been elaborated to use terminology data for customization and quality improvement of domain specific statistical machine translation. Training level integration includes enrichment of monolingual and parallel data with terminology, adaptation of translation model by adapting and filtering phrase table, and adaptation of language model. Translation level integration is provided by pre-processing the source text to identify terms and mark-up them with translation hypotheses. Evaluation results show that combination of these methods significantly improves translation quality.