# TTC: Terminology Extraction, Translation Tools and Comparable Corpora

**Seventh Framework Programme**
**Call FP7-ICT-2009-4, ICT-2009.2.2: Language-based interaction**
**Small or medium-scale focused research project (STREP)**
**Grant Agreement n° 248005**
http://www.ttc-project.eu

| List of partners |
| --- |
| University of Nantes, France (coordinator) |
| University of Stuttgart, Germany |
| University of Leeds, United Kingdom |
| Sogitec Industries, France |
| Syllabs SAS, France |
| Tilde, Latvia |
| Eurinnov, France |

**Project duration: January 2010 – January 2012**

## Summary

TTC focuses on semi-/automatic acquisition of aligned bilingual terminologies from comparable corpora, i.e. from texts of the same domain (and possibly genre) in different languages. TTC develops techniques for the collection of comparable corpora, extraction of monolingual term candidates and their contexts for English, German, French, Spanish, Latvian, Russian and Chinese from the collected corpora. Furthermore, TTC defines and combines different symbolic and statistical strategies for the alignment of the extracted monolingual term candidates. TTC develops the software for project languages and tests it on several language pairs. The tools are provided as a standalone package and web service and include components for corpora crawling and management, monolingual term candidate extraction and alignment. An integration with EuroTermBank and selected computer-assisted translation tools and machine translation systems will be provided. TTC will evaluate the quality of machine translation which can be achieved by enhancing machine translation systems with automatically extracted terminology. The main target groups of the TTC tools are professionals from the translation, localization and/or documentation industry.

TTC is at the beginning of its second year now and so far the project has made significant progress and achieved its first goals:

- Requirements and definitions of the TTC tools were specified.
- The first TTC workshop with end users was successfully held.
- A crawler was developed and initial comparable corpora for project languages were collected and analyzed.
- Term identification issues were researched for project languages and first experiments on term extraction were performed for English, German and French.