



MOLTO Enlarged EU – Multilingual On-Line Translation

EU

FP7/2007-2013

Small or medium-scale focused research project

Grant agreement No. 288317

<http://www.molto-project.eu>

List of partners
Göteborgs Universitet, Sweden (coordinator)
Helsingin yliopisto, Finland
Universitat Politècnica de Catalunya, Spain
Ontotext AD, Bulgaria
Be Informed, The Netherlands
University of Zurich, Switzerland

Project duration: March 2010 — May 2013

Summary

The MOLTO project aims to provide technology which can simultaneously tackle issues arising from real-time machine translation of web documents: localization to several languages, maintenance of their consistency in spite of asynchronous collaborative authoring with frequent edits, and grammatically and stylistically flawless text. Fifteen languages will be covered in the translations, including 12 of the 23 official languages of the European Union: Bulgarian, Danish, Dutch, English, Finnish, French, German, Italian, Polish, Romanian, Spanish, and Swedish. The 3 non-EU languages are Catalan, Norwegian, and Russian.

Two problems have slowed down the adoption of high-quality restricted language translations: development cost for a new domain or language, and learning curve for authoring texts in a restricted language. MOLTO tools will decrease the effort of developing restricted language translators radically by using the Grammatical Framework (GF) libraries. MOLTO editing tools are now available in initial prototypes for grammarians: the web-based editor and the GF plugin for the Eclipse integrated development environment. New members of the MOLTO Enlarged EU Consortium are extending the tools to a semantic wiki platform and testing the user-friendliness of these solutions for non-expert grammar writers.

MOLTO is exploring the two-way interoperability of grammars with Semantic Web conceptual models and hybrid models of combining rule-based translation systems with statistical machine translation. Data sets made available in a machine readable form, like RDF or OWL, can be used to construct a knowledge infrastructure suited to meaningful query and retrieval using natural languages. This approach is already being demonstrated with an application to the domain of cultural heritage. Combination approaches studied in MOLTO aim to integrate grammar-based and SMT models in a hybrid, robust MT system. Variants under consideration include e.g. soft integration in which phrase pairs or tree fragment pairs, generated by GF, are integrated as a discriminative probability models in a phrase-based SMT system. The testbed application for this research activity is information retrieval from patents in the pharmaceutical domain.