



Abu-MaTran: Automatic building of Machine Translation

FP7-PEOPLE-2012-IAPP

<http://www.abumatran.eu>

List of partners	
	Dublin City University, Ireland (coordinator)
	Prompsit Language Engineering SL, Spain
	Universitat d'Alacant, Spain
	University of Zagreb, Faculty of Humanities and Social Sciences, Croatia
	Athena Research and Innovation Center in Information Communication & Knowledge Technologies, Greece

Project duration: January 2013 — December 2016

Summary

Abu-MaTran seeks to enhance industry–academia cooperation as a key aspect to tackle one of Europe’s biggest challenges: multilingualism. We aim to increase the hitherto low industrial adoption of machine translation by identifying crucial cutting-edge research techniques (automatic acquisition of corpora and linguistic resources, pivot-language techniques, linguistically augmented statistical translation and diagnostic evaluation), making them suitable for commercial exploitation. We also aim to transfer back to academia the know-how of industry to make research results more robust. We work on a case study of strategic interest for Europe: machine translation for the language of a new member state (Croatian) and related languages. All the resources produced will be released as free/open-source software, resulting in effective knowledge transfer beyond the consortium. The project has a strong emphasis on dissemination, through the organisation of workshops that focus on inter-sectoral knowledge transfer. Finally, we have a comprehensive outreach plan, including the establishment of a Linguistic Olympiad in Spain, open-day activities and the participation in the Google Summer of Code.

At EAMT 2014 we will present the results of the first milestone of the project (July 2013), a general-domain MT system for English–Croatian based on free/open-source software and publicly available data, released on July 1st 2013 to mark Croatia's accession to the EU. We will also present ongoing work towards the second milestone (December 2014) including (i) a domain-specific MT system for English–Croatian in the domain of tourism, (ii) generation of synthetic English–Croatian data via Slovene using quality estimation, (iii) outcomes of the first edition of the Linguistic Olympiad of Spain (September 2013 - March 2014), etc.