



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 2015 we will present the results of the second milestone of the project (December 2014). To mention just a few: (i) MT systems for English–Croatian based on free/open-source software and web crawled and publicly available data, both generic and specific for the tourism domain, (ii) tools developed in the project (e.g. web crawling of parallel data and paradigm guessing) and (iii) outcomes of the project's dissemination activities (e.g. software management for researchers, data creation for RBMT systems and establishment of a linguistics Olympiad).