Abu-MaTran: Automatic building of Machine Translation

Antonio TORAL[1], Sergio ORTIZ_ROJAS[2], Mikel FORCADA[3], Nikola LJUBESIC[4], Prokopis PROKOPIDIS[5]

[1]ADAPT Centre, School of Computing, Dublin City University, Ireland
[2]Prompsit Language Engineering SL, Spain
[3]Departament de Llenguatges i Sistemes Informatics, Universitat d'Alacant, Spain
[4]Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
[5]Athena Research and Innovation Center, Greece

atoral@computing.dcu.ie

Abstract. We present the current status of Abu-MaTran (http://www.abumatran.eu), a 4-year project (January 2013–December 2016) on rapid development of machine translation for under-resourced languages. It is funded under Marie Curie's Industry-Academia Partnerships and Pathways 2012 programme. This is a consortium-based project with 5 partners (4 academic and 1 industrial).

Description

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 (MT) by identifying crucial cutting-edge research techniques, 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: MT for the language of a new member state (Croatian) and related languages. All the resources produced are released as free/open-source software, resulting in effective knowledge transfer beyond the consortium.

At EAMT 2016 we will present a selection of the latests developments of the project: (i) state-of-the-art statistical and rule-based MT systems for South-Slavic languages based on free/open-source software, web crawled and publicly available data and linguistic knowledge, (ii) a novel tool for massive crawling of parallel and monolingual data from the Internet's top level domains and (iii) outcomes of the project's transfer and dissemination activities, e.g. MT hybridisation and web crawling for industry uses, rapid data creation for rule-based MT systems and establishment of a national linguistics Olympiad. All the resources developed within the project are freely available[1] and the MT systems deployed can be tested online.[2]

---

[1] http://www.abumatran.eu/?page_id=351

[2] http://translator.abumatran.eu/