

A half-day workshop at LREC 2014
Tuesday, 27 May 2014.
Reykjavik (Iceland)

SALTMIL: <http://ixa2.si.ehu.es/saltmil/>
LREC 2014: <http://lrec2014.lrec-conf.org/en/>
Website: <http://ixa2.si.ehu.es/saltmil/>

Registration: <http://lrec2014.lrec-conf.org/en/registration/>

Workshop Programme **([Proceedings](#))**

09:00 – 09:30 Welcoming address by Workshop co-chair Mikel L. Forcada

09:30 – 10:30 Oral papers

Iñaki Alegria, Unai Cabezón, Unai Fernández de Betoño, Gorka Labaka, Aingeru Mayor, Kepa Sarasola and Arkaitz Zubiaga
Wikipedia and Machine Translation: killing two birds with one stone

Gideon Kotzé and Friedel Wolff
Experiments with syllable-based English-Zulu alignment

10:30 – 11:00 Coffee break

11:00 – 13:00 Oral papers

Inari Listenmaa and Kaarel Kaljurand

Computational Estonian Grammar in Grammatical Framework

Matthew Marting and Kevin Unhammer

FST Trimming: Ending Dictionary Redundancy in Apertium

Hrvoje Peradin, Filip Petkovski and Francis Tyers

Shallow-transfer rule-based machine translation for the Western group of South Slavic languages

Alex Rudnick, Annette Rios Gonzales and Michael Gasser

Enhancing a Rule-Based MT System with Cross-Lingual WSD

13:00 – 13:30 General discussion

13:30 Closing

Workshop Organizers

Mikel L. Forcada Universitat d'Alacant, Spain

Kepa Sarasola. Euskal Herriko Unibertsitatea, Spain

Francis M. Tyers. UiT Norgga árkálaš universitehta, Norway

Workshop Programme Committee

Iñaki Alegria Euskal Herriko Unibertsitatea, Spain

Lars Borin Göteborgs Universitet, Sweden

Elaine Uí Dhonnchadha Trinity College Dublin, Ireland
Mikel L. Forcada Universitat d’Alacant, Spain
Michael Gasser Indiana University, USA
Måns Huldén Helsingin Yliopisto, Finland
Krister Lindén Helsingin Yliopisto, Finland
Nikola Ljubešić Sveučilište u Zagrebu, Croatia
Lluís Padró Universitat Politècnica de Catalunya, Spain
Juan Antonio Pérez-Ortiz Universitat d’Alacant, Spain
Felipe Sánchez-Martínez Universitat d’Alacant, Spain
Kepa Sarasola Euskal Herriko Unibertsitatea, Spain
Kevin P. Scannell Saint Louis University, USA
Antonio Toral Dublin City University, Ireland
Trond Trosterud UiT Norgga árktalaš universitehta, Norway
Francis M. Tyers UiT Norgga árktalaš universitehta, Norway

Introduction

The 9th International Workshop of the Special Interest Group on Speech and Language Technology for Minority Languages (SaLTMiL) will be held in Reykjavík, Iceland, on 27th May 2014, as part of the 2014 International Language Resources and Evaluation Conference (LREC). (For SaLTMiL see: <http://ixa2.si.ehu.es/saltil/>); it is also framed as one of the activities of European project Abu-Matran (<http://www.abumatran.eu>). Entitled "Free/open-source language resources for the machine translation of less-resourced languages", the workshop is intended to continue the series of SaLTMiL/LREC workshops on computational language resources for minority languages, held in Granada (1998), Athens (2000), Las Palmas de Gran Canaria (2002), Lisbon (2004), Genoa (2006), Marrakech (2008), La Valetta (2010) and Istanbul (2012), and is also expected to attract the audience of Free Rule-Based Machine Translation workshops (2009, 2011, 2012).

The workshop aims to share information on language resources, tools and best practice, to save isolated researchers from starting from scratch when building machine translation for a less-resourced language. An important aspect will be the strengthening of the free/open-source language resources community, which can minimize duplication of effort and optimize development and adoption, in line with the LREC 2014 hot topic ‘LRs in the Collaborative Age’ (<http://is.gd/LREChot>).

Papers describe research and development in the following areas:

- Free/open-source language resources for rule-based machine translation (dictionaries, rule sets)
- Free/open-source language resources for statistical machine translation (corpora)
- Free/open-source tools to annotate, clean, preprocess, convert, etc. language resources for machine translation

- Machine translation as a tool for creating or enriching free/open-source language resources for less-resourced languages