Host Report Short Term Scientific Mission IS1305- 210914-049230

COST STSM Reference Number: COST-STSM-IS1305- 210914-049230
Period: 2014-09-21 00:00:00 to 2014-09-27 00:00:00
COST Action: IS1305

STSM type: Regular (from Slovakia to Slovenia)
Guest/STSM applicant: Vladimir Benko, Slovak Academy of Sciences, L. Stur Institute of Linguistics, Slovakia
Host: Dr. Iztok Kosem, Trojina, Institute for Applied Slovene Studies iztok.kosem@trojina.si
Duration: 7 working days : from 21st September 2014 till 27th September 2014.

The Scientific mission had several aims:
   a) to devise the first version of GDEX for Slovak, based on the GDEX for Slovene, and applying it to the Slovak corpus in the Aranea family.
   b) to include the Slovene corpus to the Aranea family
   c) to conduct initial testing of bilingual word sketches.

The applicant was given access to Slovenian corpora that are open access, and enabled to upload them to the server holding Aranea family corpora. Based on this, the applicant tested bilingual word sketches for English-Slovene, which were evaluated by dr. Kosem. The plan was made to prepare a joint publication that will evaluate bilingual word sketches and consider their lexicographic potential.

The applicant was also provided with GDEX for Slovene, version 1, along with descriptions of various classifiers. Based on this, the applicant produced the first version of GDEX for Slovak, and tested it on the corpus data. This required checking and modifying language non-specific classifiers, and removing or localizing the language specific classifiers.

The applicant also met with several linguists and language technologists from other institutions that cooperate with Trojina, such as Faculty of Arts, University of Ljubljana and "Jožef Stefan" Institute in Ljubljana. Furthermore, during his visit, the applicant attended a workshop on automatic data extraction, organized by Consortium for Language Resources and Technologies (Konzorcij za jezikovne vire in tehnologije), of which Trojina is also a member. At the workshop, the applicant got familiar with state-of-the-art of language data extraction in Slovenia, including definition extraction, extraction of collocations and good dictionary examples, etc.

Finally, the applicant was also invited to give a talk on Digitization of Historical dictionaries of Slovak at the ENeL workshop in Bled (29-30 September 2014).