REPUBLIC OF MACEDONIA
Ss. Cyril and Methodius University in Skopje
Faculty of Computer Science and Engineering

P.O. Box 393, 1000 Skopje; Tel: 02/30 88 255; www.finki.ukim.mk

Final project proposal

Type	Undergraduate or master
Title	***Bilingual phraseological dictionary***
Supervisor	Katerina Zdravkova
e-mail	katerina.zdravkova@finki.ukim.mk
Department / Group	
Institute of Intelligent Systems	
Topic(s)	
Natural Language Processing, Machine Learning	
Project can start from	01.09.2014 – 31.12.2014
Project duration	4 months
Short description	
One of the greatest challenges of bilingual translation is to find the most accurate target phrase, idiom, or syntagm, whenever it exists. The main goal of this project is to predict the prospective target multi-word expressions (MWEs) in the student's native language (or English) based on the corpus of Orwell's 1984. The initial MWEs will be extracted from the Macedonian phraseological dictionary and from Orwell's novel. They exist in the exact or in a similar form in the aligned multilingual Multext-East corpus. Using machine learning techniques, target phrases have to be predicted and evaluated with the Multext-East version in student's native language. The same approach should be implemented in the opposite direction, from target to source language and re-evaluated. The project can be performed by one student or by several students from different countries, enabling cross-evaluation of obtained results.	
Results and assessment	
Student is supposed to develop a bilingual dictionary system that inserts any bilingual document, extracts its MWEs, predicts the target MWEs, and cross-evaluate using the opposite direction. The creation of the system, as well the results of its evaluation should be reported and publicly presented. The efficiency of the system and the quality of the final report will be the crucial elements assessed by the supervisor.	
Other (additional) information	
Basic resources are available from Multext-East project (http://nl.ijs.si/ME/). The project fits with the Parseme COST action (http://typo.uni-konstanz.de/parseme/)	