

Project ID:
Reporting period:

7F14047
01. 01. 2016 – 30. 04. 2017

Annex V - Evaluation Report of the Project

Mandatory outline of opponent's review (for writing the evaluation, please, use the headings 1-6).

1. Evaluation of project periodic report

Project Periodic Report for period 1 January 2016 to 30 April 2017 evaluates the fulfillment of the HaBiT project during the above cited period. I certify that all proposed work packages were fulfilled during the evaluation period, several results of the project were already presented at scientific conferences, a comprehensive list with detailed descriptions is stated in the section 2.8.1 of the Project periodic report. The main aim of the project accomplishment was to make all project outputs open access, the resources as well as the tools. Main efforts of all team members were devoted to analysis and implementation of new techniques and methods for gathering and shallow linguistic processing of new large corpora of the specialized languages, mainly for major Ethiopian languages. The main results of the Project for the evaluation period are the following:

- to gather large-scale text data (corpora) for under-resourced languages from the Web and process them so they can be used in language applications, such as information extraction or machine translation; these languages include Norwegian, Czech and four major Ethiopian languages: Amharic, Afaan Oromo, Tigrinya and Somali;
- to build a multi-billion word Norwegian corpus using the tools co-developed by Masaryk University and utilized in a joint EU-funded project with NTNU ("PRESEMT: Pattern REcognition-based Statistically Enhanced MT", 2010-2012);
- to feed into and leverage on the NORHED project ("Linguistic Capacity Building – tools for the inclusive development of Ethiopia", NORHED 2013-2018, collaboration of NTNU, University of Oslo and two Ethiopian universities) and include the processing of the Ethiopian languages mentioned above into the HaBiT project;
- to build shallow processing applications for Czech, Norwegian and Amharic for investigating and separating multiple senses of the words in the corpora, as well as for creating multi-sense vector spaces and parallel multi-lingual vector spaces for word translation disambiguation.

2. Evaluation of activities, outputs and objectives of project set in the decision on the provision of the institutional support including its annexes (hereinafter referred to as "Decision")

Seven project work packages were successfully fulfilled, the presented results of the project are valuable, they help to all cited participants, mainly in Africa, substantially improve the communication by major languages of Ethiopia and they enable to develop and to build shallow processing applications. The achieved results of the project are clearly presented, the presentation is readable without problems. The realization of the project is based on the right base and it seems me to be without several troubles (from the report).

Several extensions of corpora for four African languages and a Czech-Norwegian parallel corpus were developed within the framework of this project. Large new corpora (in case of the Ethiopian languages the world's-largest) have been prepared that can be considered as one of main results of this project. The corpora are accompanied by several software products that were developed to build new corpora and for easy manipulation of existing corpora, esp. for easy use by novice users. The use of software is clear and easy and it should be a simple matter to extend them for other developed corpora. The scientific part's schedule is kept without comments, the cooperation of the project partners can be evaluated as successful, and the project program was fulfilled completely. Utilization and exploitation of results of the project is be relatively wide, mainly in Africa, and all results contribute to further development of the language technology in the future.

3. Evaluation of scientific realization of project

The scientific level of the project is high, professional quality of team too. The expected and also already achieved results bring several new scientific results – it was already confirmed by many presented publications. The project contributes to research-based knowledge development by extending the cooperation within and outside of the Czech-Norwegian Research Programme. A number of PhD students are already part of the research and the project partners plan to include more female researchers, PhD students, and postdocs in order to actively engage them in best practices of both workplaces, and to transfer and share state-of-the-art development. This close cooperation between the partners becomes a basis for future successful cooperation, thus ensuring sustainability of the research or other future relevant activities. The project results also contribute to the cultural development of less-developed countries by acquiring and sharing developed information technologies. All project outputs were presented to the international research community and are also disseminated via the project website in order to present the research to all possible types of audiences.

4. Assessment of actual incurred expenditure of project

The incurred expenditure was utilized reasonable, the financial means (money amounts) can be evaluated as appropriate and their utilization by all participants of the project rational. The use of financial means is good presented in the report (on pages 35 – 37). There are presented only small operative changes in the part of the budget planned for 2016 – 17 and these changes do not influence the MU total project budget, neither the total personnel expenses. No procurement or small scale contracting were planned for the project, so none were launched in the reporting period.

5. Requirements on project report modification

There were no requirements for project modification.

6. Final evaluation of project

The project has been managed during the reporting period in a standard way. All presented results are valuable and they contribute to scientific progress in language technology. They bring new scientific information/pieces of knowledge and the plan of research actions is kept without substantive changes. Altogether 31 scientific publications are cited in the report. The project succeeded in creation of a repository for the investigated languages, making them accessible for further research, enabling to acquire information technologies in a less-developed countries (such as Ethiopia) and thus contributing to its cultural development. The accessibility of the results contributes to further development of research in the area of under-resourced languages. Number of PhD students and postdocs were actively participating in research, helping to transfer and share state-of-the-art development. The close cooperation between partners led to numerous project outputs that were presented to the international research community and have been disseminated via the project web page aiming at presenting the research to any kind of audience. The financial means were drawn reasonable and with small changes in budget, personnel changes did not affect the project budget. I recommend to close the project solving with a great success.

Surname, first name, degree	Matoušek Václav, Prof. Ing., CSc.
Work place:	University of West Bohemia in Plzeň (Pilsen)
Statement of opponent: I hereby declare that I have been given all the necessary information about the project and all the required documents for assessing of this project from the principal investigator, and I was solely driven by objective considerations.	
Date: 13. 06. 2017	(e-)Signature: 