

**Identifikační kód projektu/Project ID:** 7F14047  
**Doba implementace/Implementation period:** 1. 10. 2014 – 30. 4. 2017

## **Závěrečný oponentní posudek** **Final Evaluation Report**

Povinná osnova závěrečného oponentního posudku (tj. body 1-5). Posudek k výzkumnému projektu musí být vypracován **v anglickém jazyce**/The mandatory outline of the final evaluation report of the project (points 1-5). The report must be filled out in English.

### **1. Zhodnocení dosažení plánovaných cílů a výsledků stanovených v právním aktu v době řešení projektu/Evaluation of achievement of the planned objectives and outcomes set in the project contract during the project implementation.**

The main objectives of the research project No. 7F14047 named HaBiT were:

- 1) 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;
- 2) 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;
- 3) 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;
- 4) 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.

The aims above were accomplished in delivering the final HaBiT system and addressing the topics on technology assessment, verification and testing, as well as on ICT meeting societal challenges, hence obtaining a relevant added value also in the political respect through cooperation with a less-developed country.

All proposed work packages were fulfilled during the project solving period, several results of the project were already presented at scientific conferences, a comprehensive list with detailed descriptions is stated in the section 2.5 of the Final Project Report. The main efforts were devoted to analysis and implementation of new techniques and methods for gathering and shallow linguistic processing of new large corpora of the respective languages. Large annotated corpora were created for all given languages and software modules, such as taggers, parsers and sketch grammars were created. The results were presented to the scientific community via conference and journal papers as well as the project web page. The outputs are freely accessible for further research via the HaBiT system created with cooperation of project partners (Masaryk University, Norwegian University of Technology, University of Oslo, and two Ethiopian universities, which cooperate with NTNU on NORHED project. The HaBiT project leveraged on the NORHED project, thoroughly testing the technologies and thus addressing the call topics on technology assessment, verification and testing, as well as on ICT meeting societal challenges, hence obtaining a relevant added value also in the political respect through cooperation with a less-developed country.

### **2. Zhodnocení kvality vědeckého publikování vzešlého z projektu/Evaluation of the scientific publications originated from the project.**

☒ **5 Outstanding** ☐ **4 Very good** ☐ **3 Good** ☐ **2 Satisfactory** ☐ **1 Unsatisfactory**

There are three kinds of publications – the first one is the information about the project. This information is presented online on the project web page (<http://habit-project.eu/>) and also on the Masaryk University Web page. The second one are the publications in journals and proceedings –

there are totally 47 high quality publications, three of them were published in internationally referred (joint) scientific journals. And as the third kind of publications can be seen the special presentations on conferences and sessions and their list is completely published in part 2.5 of the final report. These tools were used to target research community and students as well as broader public.

**3. Zhodnocení Závěrečné zprávy o řešení projektu a požadavky na úpravy či doplnění Závěrečné zprávy o řešení projektu/Evaluation of the Final Project Report and requirements on modifications and completion of the Final Project Report.**

The Final Project Report is written very responsibly, it contains all necessary parts and all parts of the report are good understandable. The outputs of the project fully cited in the report are freely accessible for further research via the HaBiT system created with cooperation of other project partners. It is not necessary to modify the report.

**4. Závěrečné celkové hodnocení celého projektu včetně nesouhlasných názorů/Final overall evaluation of the whole project including dissenting views.**

☒ 5 Outstanding ☐ 4 Very good ☐ 3 Good ☐ 2 Satisfactory ☐ 1 Unsatisfactory

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 47 scientific publications and presentations are cited in the final report. The project status did not undergo any change, no deviations from the planned milestones have occurred, and all deliverables have been prepared on time, only with the exception of Deliverable 1.1.4 which have been postponed awaiting the employment of a further postdoc researcher at NTNU. The financial means were drawn very reasonable and with small changes in budget, personnel changes did not affect the project budget. The funding of the project corresponded in case of both the Project Promoter and the Project Partner to fast 100% of the planned project budget, the remaining amount (see the pages 12 – 13 of the final report) is going to be returned to the Project Promoter. This amount corresponds to 2% of the planned expenditures, which was caused by savings in the personnel and travel expenses.

**5. Byl projekt dokončen podle právního aktu? Vyberte jednu z možností Ano, či Ne/Was the project completed according to the project contract? Choose Yes, or No.**

☒ ANO/YES

No comment to this part, the project was fully completed according to the project contract.

**Příjmení, jméno, titul oponenta/Surname, first name, degree of the opponent:**

Matoušek, Václav, Prof. Ing., CSc.

**Pracoviště/ University of West Bohemia in Plzeň (Pilsen)**

**Čestné prohlášení oponenta/Statement of the opponent:**

**Čestně prohlašuji, že mi byly řešitelem projektu včas poskytnuty všechny potřebné informace a předloženy mnou požadované doklady a při posuzování tohoto projektu jsem se řídil(a) výhradně objektivními hledisky./I hereby declare that I have been given all the necessary information about the project and all the required documents for assessment of this project from the principal investigator, and I was solely driven by objective considerations.**

**Datum/Date:** 15.06.2017

**Místo/Place:** Plzeň



**(e-)Podpis/(e-)Signature:**