

AGREED MINUTES ON OUTCOME OF FINAL EXTERNAL EXAMINATION

Czech-Norwegian Research Programme CZ09

Norway Grants 2009 – 2014

A. PROJECT INFORMATION

Project ID	7F14047			
	from	01. 10. 2014	to	30. 04. 2017
	Harvesting big text data for under-resourced languages			
	prof. PhDr. Karel Pala, CSc.			
	25 593 000			

Project promoter	Masarykova univerzita	<i>research organization</i>
Project partner	Norges teknisk-naturvitenskapelige universitet	<i>research organization</i>

**Choose one of these: research organization/or small enterprise/or medium enterprise.*

B. PEER REVIEW BOARD'S EVALUATION (to be filled out by the peer review board)

1.	Were the work packages/tasks completed in accordance with the schedule stipulated in the project contract?	<input checked="" type="checkbox"/> YES	<input type="checkbox"/> NO
<p><i>Please comment (min. 300 characters):</i></p> <p>All proposed work packages were fulfilled during the project solving period, the project has more than achieved its goals as set out in the original project proposal and in the proposal for additional research activities. 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. The plan of research actions was kept without substantive changes. The change to named entity recognition from word sense disambiguation was sensible given the data available, changes in the state-of-the-art in computational linguistics and the overriding goal of semantic search.</p> <p>It is possible to certify that the project was fulfilled completely in time.</p>			
2.	Were the planned deliverables/milestones/outcomes stipulated in the project contract achieved?	<input checked="" type="checkbox"/> YES	<input type="checkbox"/> NO
<p><i>Please comment (min. 300 characters):</i></p> <p>All planned deliverables (in total 31) were successfully fulfilled, all planned outputs were delivered as well: annotated corpora and sketch grammars for all involved languages were created, evaluation methodology for sketch grammar output, named entity recognition software, interactive visualizations of the data and the PoS annotation framework was developed. Therefore the HaBiT system was completely finished. The planned aims 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. The expected outcomes of the project are large annotated corpora for the participating under-resourced languages and Part-of-Speech annotation framework as well as taggers, parsers and sketch grammars for the involved languages. All the outcomes were achieved as planned and are accessible via the publicly available interface of the HaBiT system.</p>			
3.	Were the project objectives achieved?	<input checked="" type="checkbox"/> YES	<input type="checkbox"/> NO
<p><i>Please comment (min. 300 characters):</i></p> <p>The objective of the project was to create annotated corpora, annotation framework and to help acquiring information technology in a less-developed country. This objective was fulfilled in both aspects:</p>			

<p>a) general methodologies for obtaining linguistic resources (corpora, taggers, sketch grammars) for a new language were developed and published, and</p> <p>b) the methodologies were explicated in the form of the respective linguistic resources for the four Ethiopian languages, where they represent the state-of-the-art for these languages.</p> <p>This objective was completely fulfilled and all results of it 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).</p>		
4.	Was dissemination of research outcomes optimal concerning the effective promotion of the project?	<input checked="" type="checkbox"/> YES <input type="checkbox"/> NO
<p><i>Please comment (min. 300 characters):</i></p> <p>The dissemination of all of planned outputs can be seen as optimal - the publication output of the project (altogether 47 scientific publications and presentations cited in the final project report) is of high quality as well as prolific (well in excess of one publication per project month). It includes a number of high-profile international conference publications (ACL conferences and workshops, RANLP, TSD), a couple of articles in respected journals, and also contributions to conferences and workshops focusing on particular under-resourced languages. Only one possible reservation concerns integration. Looking through the publication output of the project, it can be noted a distinct lack of publications with co-authors from more than one of the collaborating groups. Both the MU group and the NTNU group have produced high-quality scientific outputs that in all cases conform to and promote the project goals as set out in the project proposal.</p> <p>However, a collaborative project of this kind should ideally exhibit a greater degree of integration in its outputs, and the lack of this mars the overall impression somewhat. This is something that the project partners may want to consider in planning future collaborations.</p>		
5.	Did the Project promoter apply the substantive recommendations and/or measures from the performed on-the-spot audit(s) during the project implementation?¹	<input type="checkbox"/> YES <input type="checkbox"/> NO <input checked="" type="checkbox"/> NOT APPLICABLE
<p><i>Please comment (min. 300 characters):</i></p> <p>The project passed an on-the-spot audit by an independent auditor INTEREXPERT BOHEMIA spol.s r.o. with the audited period of Oct 1, 2014 to Dec 31, 2016. The result of the audit was <i>Correct implementation</i> and the auditor certified that</p> <p>a. <i>claimed actual costs within the project implementation are incurred in accordance with the Regulation on the implementation of Norwegian Financial Mechanism 2009-2014.</i></p> <p>b. <i>claimed actual costs within the project implementation are incurred in accordance with the national law and accounting practices of the entity country.</i></p>		
6.	Importance and sustainability of the achieved project outcomes: <ul style="list-style-type: none"> Scientific excellence Economic and/or societal usability of the project outcomes and their sustainability 	
<p><i>Please comment (min. 300 characters):</i></p> <p>Intelligent language processing applications (such as information extraction or machine translation) need processed (annotated and parsed) large text data. Big languages (English, German, ...) have enough resources available, however, less-covered languages such as the Ethiopian languages but also hundreds of other languages, are in urgent need of methodologies and techniques for linguistic resource creation. The HaBiT project developed such methodologies and techniques and presented them with the case of four main Ethiopian languages.</p> <p>The financial means were drawn very reasonably 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 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.</p> <p>The corpora produced by the project are made openly available on the project page, however, it would greatly increase the sustainability and visibility of the data if they were also made available via the repository of the Czech node of CLARIN.</p> <p>The two international workshops organized within the HaBiT project received high ratio of attendants outside the project partners from all over the world. The interim results of the project have been published in 40 scientific conference papers and two international journal articles. The application results are publicly available in the form of software tools and linguistic databases.</p>		

C. PEER REVIEW BOARD'S OVERALL CONCLUSION (Please choose one option of i/ii/iii overall conclusions only)

i. THE PROJECT CONTRACT MAY BE DEEMED AS EXECUTED <i>If YES, please comment and explain.</i>		<input checked="" type="checkbox"/> YES
<i>Please comment (min. 200 characters):</i> The project has much more than achieved its goals as set out in the original project proposal and in the proposal for additional research activities. All presented results are valuable and they contribute to scientific progress in language technology and bring new scientific information/pieces of knowledge. The plan of research actions was kept without substantive changes, the financial means were drawn very reasonable and with very small changes in budget, personnel changes did not affect the project budget.		
ii. THE PROJECT CONTRACT MAY BE DEEMED AS EXECUTED PARTIALLY <i>If YES, in this case please state what part of the project contract were not performed such as a number of a work package, a milestone, an outcome etc. or specify the percentage of non-performance of the project. Please state particular reasons for incompleteness of the project.</i>		<input type="checkbox"/> YES
<i>Please comment (min. 300 characters):</i>		
iii. THE PROJECT CONTRACT MAY BE DEEMED AS NON-EXECUTED <i>If YES, in this case please state particular reasons for incompleteness of the project.</i>		<input type="checkbox"/> YES
<i>Please comment (min. 300 characters):</i>		
SCIENTIFIC LEVEL OF THE PUBLICATIONS ORIGINATED FROM THE PROJECT (Please mark rating from 5 to 1) <input type="checkbox"/> 5 Outstanding <input checked="" type="checkbox"/> 4 Very good <input type="checkbox"/> 3 Good <input type="checkbox"/> 2 Satisfactory <input type="checkbox"/> 1 Unsatisfactory		
<i>Please comment (min. 300 characters):</i> The project produced a large number of conference publications as well as some journal publications, both very good for the scope of the evaluated project. The project could have more publications jointly authored by the partners of the project, as well as having a journal publication summarising the complete project.		
OVERALL PROJECT RATING (Please mark rating from 5 to 1) <input checked="" type="checkbox"/> 5 Outstanding <input type="checkbox"/> 4 Very good <input type="checkbox"/> 3 Good <input type="checkbox"/> 2 Satisfactory <input type="checkbox"/> 1 Unsatisfactory		
<i>Please comment (min. 300 characters):</i> The project produced an impressive number of concrete results, in particular the corpora of the African languages, which are both downloadable as well as searchable via an installed Sketch Engine interface. The project delivered all the major promised results, and, in some cases, even exceeding the planned size of the datasets. For the few results not delivered, a good justification was made as to why to substitute them with other deliverables.		
Chairperson and members of the peer review board:		
Place: Skype meeting Date: June 27, 2017		
Full name	Employer	Signature
dr. Tomaž Erjavec (chair)	Jožef Stefan Institute, Slovenia	
prof. Václav Matoušek	Západočeská univerzita, Plzeň	
dr. Diana McCarthy	University of Cambridge, UK	

Signature of the Principal investigator:

prof. PhDr. Karel Pala, CSc.

Date: June 27, 2017

Place: Skype meeting