

Live minutes

NLPL steering group meeting

Time: 2017-08-28, 13:30-15:00 CEST

Place: video meeting via Zoom, <https://zoom.us/j/3989300056>

Invited:

Tomasz Malkiewicz, NeIC (PO)

Joakim Nivre, Uppsala University

Jörg Tiedemann, University of Helsinki

Martin Matthiesen, CSC-IT Center for Science Ltd

Stephan Oepen, University of Oslo

Anders Søgaard, University of Copenhagen

Filip Ginter, University of Turku

Gunnar Bøe, UNINETT Sigma2 AS

Bjørn Lindi, NeIC (PM)

Known absences: Anders Søgaard, Gunnar Bøe (will be delayed)

Unknown absences: Filip Ginter

Agenda:

NLPL SG 17-27. Attendance and agenda (5')

NLPL SG 17-28. New PM (10')

NLPL SG 17-29. Status of the project (15')

NLPL SG 17-30. Review of personnel situation (10')

NLPL SG 17-31. Project deliveries (10')

NLPL SG 17-32. Status report to NeIC board (10')

NLPL SG 17-33. Resource allocation principles (15')

NLPL AG 17-34. SG F2F meeting (10')

NLPL SG 17-35. AOB (5')

Minutes

NLPL SG 17-27. Attendance and agenda (5')

NLPL SG 17-28. New PM (10')

Bjørn Lindi working at the NTNU HPC-section since 2004. Experience from PRACE. Involved in the NeIC-projects CodeRefinery and Dellinger, as well. Started as NLPL project manager on 1.08.2017.

NLPL SG 17-29. Status of the project (15')

Activity A – Project Management and Infrastructure

Bjørn started 1. August as PM for the project.

Stephan has (finally) submitted the Service Agreement for UiO
Service Agreement for University of Copenhagen is still missing.

Project Groups has met through Slack. We (Tomasz&&Bjørn) propose to move to video meetings. It is challenge to a establish a common ground through Slack. Slack was chosen as an experiment during the Kickoff. To get all who are involved in the project to meet on Google Hangout will also be a challenge. Google Hangout has a maximum 10 participants for free, and maximum 25 participants for customers.

Stephan suggest to reduce meeting frequency for full team and rather establish task forces for (at least the) following tasks:

- technical infrastructure
- resource allocations
- outreach and training

Proposal/Decisions:

- Bjørn will for each task have a video meeting for discussing progress and how to arrange the work in the task.
- Bjørn will make a project plan based on the current work plan

Activity B – Statistical and Neural Machine Translation

D B.1.1 Moses release 3.0 and support tools is **done**

Activity C – Dependency Parsing

All software is available on Taito. Need to provide installation of parsing software on Abel, where this activity was intended to have its primary home. Original challenges installing TensorFlow and DyNet on Abel appear resolved.

Activity D – (Very) Large Corpora

Some resources are under licenses. Many partners in NLPL has licenses. Licensing to the laboratory, and not individual partners has been considered and investigated. This will not happen. The Project will need to manage access if licensed resources are put on the a shared system. We should Reassess the need for the licensed data. A few licensed data sets might be somewhat critical. Worth considering a position where NLPL provides only free resources, and each partner (or user) is free to bring their own licensed data. REMS comes in a new version. Task D2 Corpora-provisioning is ahead of schedule.

D1.1 is **done**, obtaining an LDC license for NLPL at large is not possible.

D1.2 is not worth the effort and is **undone**.

Activity E – Embeddings

Task E.1 is complete, but documentation is missing; see <http://hdl.handle.net/11234/1-1989> and <http://vectors.nlpl.eu>. More and more embeddings will happen before M24, Possibly invite test users from within the consortium.

D E.1 Pre-trained embeddings for ENG, DAN, FIN, NNO, NOB SWE is **done**.

Activity F – Extrinsic Evaluation

Copenhagen the coordinator, but so far mostly Oslo, Turku, and Uppsala have worked on this. Three downstream systems with a uniform interface are available. The infrastructure has already been used by people from outside the consortium. see <http://epe.nlpl.eu>

Activity G – Parallel Corpora and OPUS

The OPUS server is available. Mirroring the OPUS data will probably not happen due to space limitation on Abel. Possible mirror NorStore, but data will then be offline, and need to be checked out/transferred before use.

D G.1.1 Running OPUS server is **done**

D G.1.2 Mirroring of OPUS data is not done.

D G.2.1 Tools and documentation [M12] is mostly done, but the documentation missing. Where should documentation go?

Activity H – Community Building

The coordination and planning for Winter School 2018 must start as soon as possible. The project must decide on who to target. Should it be open to everybody or only to selected groups. The NelC Allhands-meeting is in the last week of January 2018. It is suggested to have the winter school in the continuation of the Allhands-meeting.

The program for SINTEF Winter School 2017:

<https://www.sintef.no/projectweb/geilowinterschool/2017-machine-learning/#/>

A minimal webpage exists and the project were presented with a paper and flyer on NoDaLiDa 2017. **D H.2.1** Web site [M6] is **done**.

SG 17-30. Review of personnel situation (10')

University of Copenhagen will have people available for working on the project next summer. There will be no funding from DeIC.

Separate meeting with Anders to be arranged.

Another consortium member from DK to be discussed further.

NLPL SG 17-31. Project deliveries (10')

NLPL SG 17-32. Status report to NeIC board (10')

University of Copenhagen has reported that they currently do not have personnel who can contribute to the project. At earliest UoC can work on the project from next summer - one year after schedule:

time - green

cost - yellow (Copenhagen resources not allocated)

goals - green

NLPL SG 17-33 Resource allocation principles

Resource allocation tasks force to decide and report on the usage to the SG. Martin is retroactively billing work done in the spring. Joakim has applied for more hours at CSC.

NLPL SG 17-34. SG F2F meeting (10')

Last week of October and two first of November.

Stockholm proposed as a location, Arlanda alternatively Oslo airport.

Proposal/Decision/Action:

- Bjørn will send a Doodle for an SG F2F

- Bjørn will send a Doodle for the next SG meeting, possibly end of September or beginning of October.

NLPL SG 17-35. AOB (5')

No other business