

NLPL

Stephan Oepen (oe@ifi.uio.no)

NLPL

Nordic Language Processing Laboratory

Stephan Oepen (oe@ifi.uio.no)





Network of language technology researchers in Northern Europe;
six university research groups (Denmark, Finland, Sweden, Norway);
national e-infrastructure providers in Finland and Norway;
allocations on Abel and Taito; discipline-specific software & data;
funding from NeIC, matching in-kind contributions from all partners.

So, What's in it for me?

Collaboration Infrastructure

- Distributed team of 25 or so (very) part-timers; mostly a self-help initiative;
- cross-border sharing: everyone can get access to same two superclusters;
- HPC best practices: teaching each other, and also the general support staff.



So, What's in it for me?

Collaboration Infrastructure

- Distributed team of 25 or so (very) part-timers; mostly a self-help initiative;
- cross-border sharing: everyone can get access to same two superclusters;
- HPC best practices: teaching each other, and also the general support staff.

Virtual Laboratory

- Community-maintained repository of discipline-specific software and data;
- modularity, interoperability, uniformity, reproducibility: `modules` setup;
- common (large) data sets: corpora, embeddings, parsing, translation, ...



So, What's in it for me?

Collaboration Infrastructure

- Distributed team of 25 or so (very) part-timers; mostly a self-help initiative;
- cross-border sharing: everyone can get access to same two superclusters;
- HPC best practices: teaching each other, and also the general support staff.

Virtual Laboratory

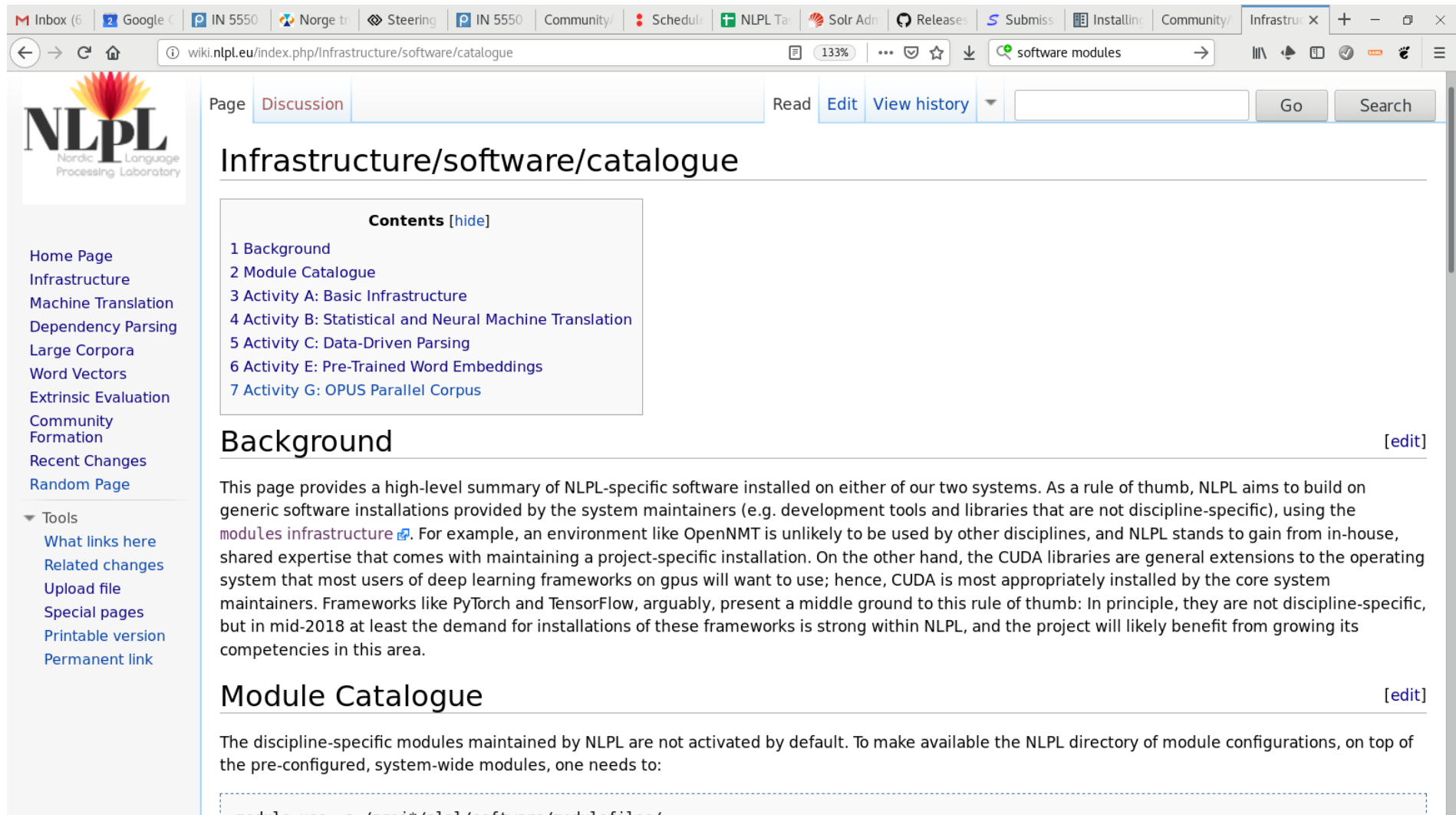
- Community-maintained repository of discipline-specific software and data;
- modularity, interoperability, uniformity, reproducibility: `modules` setup;
- common (large) data sets: corpora, embeddings, parsing, translation, ...

Meeting Place

- Kick-off meeting (2017); Annual winter school; maybe NoDaLiDa workshop.



Emerging Community-Maintained Documentation



The screenshot shows a web browser window displaying the NLPL Infrastructure/software/catalogue page. The browser's address bar shows the URL `wiki.nlpl.eu/index.php/Infrastructure/software/catalogue`. The page features a navigation menu on the left with links such as Home Page, Infrastructure, Machine Translation, Dependency Parsing, Large Corpora, Word Vectors, Extrinsic Evaluation, Community Formation, Recent Changes, and Random Page. The main content area includes a table of contents with links to Background, Module Catalogue, and seven activities (A through G). The Background section provides a high-level summary of NLPL-specific software installed on the systems, mentioning development tools and libraries like OpenNMT and CUDA. The Module Catalogue section discusses discipline-specific modules and their activation.

Contents [hide]

- 1 Background
- 2 Module Catalogue
- 3 Activity A: Basic Infrastructure
- 4 Activity B: Statistical and Neural Machine Translation
- 5 Activity C: Data-Driven Parsing
- 6 Activity E: Pre-Trained Word Embeddings
- 7 Activity G: OPUS Parallel Corpus

Background [edit]

This page provides a high-level summary of NLPL-specific software installed on either of our two systems. As a rule of thumb, NLPL aims to build on generic software installations provided by the system maintainers (e.g. development tools and libraries that are not discipline-specific), using the `modules infrastructure`. For example, an environment like OpenNMT is unlikely to be used by other disciplines, and NLPL stands to gain from in-house, shared expertise that comes with maintaining a project-specific installation. On the other hand, the CUDA libraries are general extensions to the operating system that most users of deep learning frameworks on gpus will want to use; hence, CUDA is most appropriately installed by the core system maintainers. Frameworks like PyTorch and TensorFlow, arguably, present a middle ground to this rule of thumb: In principle, they are not discipline-specific, but in mid-2018 at least the demand for installations of these frameworks is strong within NLPL, and the project will likely benefit from growing its competencies in this area.

Module Catalogue [edit]

The discipline-specific modules maintained by NLPL are not activated by default. To make available the NLPL directory of module configurations, on top of the pre-configured, system-wide modules, one needs to:

```
module use _sa /proj*/nlpl/software/modulefiles/
```



Community Formation: Annual NLPL Winter Schools

