

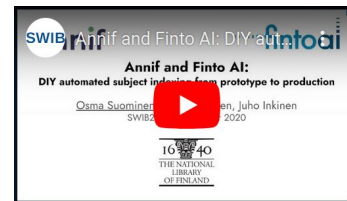
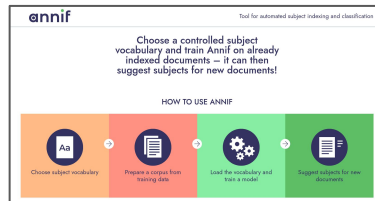
Introduction to the online hands-on tutorial



1

Understand what Annif is

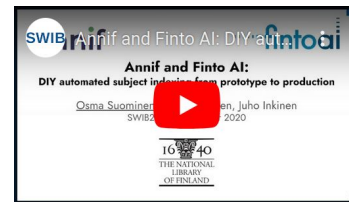
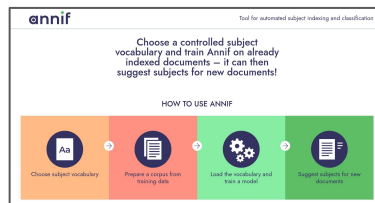
Study the website annif.org,
watch a presentation about it,
or read the J LIS.it [paper](#).



1

Understand what Annif is

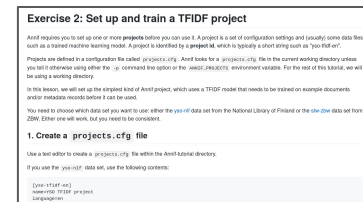
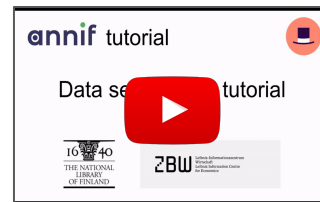
Study the website annif.org, watch a presentation about it, or read the JLIIS.it [paper](#).



2

Complete this hands-on tutorial

Watch the videos, install Annif, and complete the exercises as far as you can, on your own time.

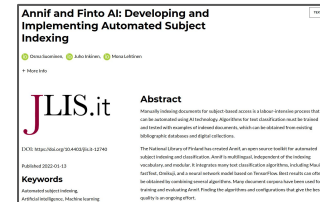
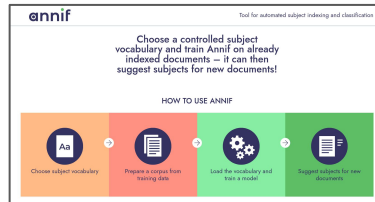


you
are
here

1

Understand what Annif is

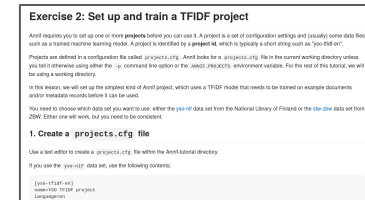
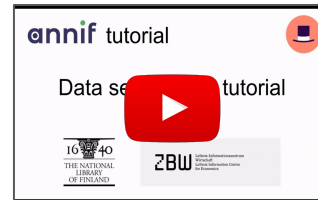
Study the website annif.org, watch a presentation about it, or read the JLIIS.it [paper](#).



2

Complete this hands-on tutorial

Watch the videos, install Annif, and complete the exercises as far as you can, on your own time.



3

Join an online session (optional)


In the online sessions, you can ask questions, get help and discuss what you've learned. Registration required.







you
are
here

Annif-tutorial GitHub repository


the main resource for this tutorial

 NatLibFi / **Annif-tutorial**

Q Type 1 to search


[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)












 **Annif-tutorial** Public





[Edit Pins](#) [Unwatch](#) 10 [Fork](#) 9 [Star](#) 36

[main](#) 2 Branches 0 Tags

[Add file](#) [Code](#)

 **jahoinkinen** Merge pull request #32 from NatLibFi/update-2024 0f858cf · 4 days ago 588 Commits








 .devcontainer	Show README.md on codespace startup (#29)	4 months ago
 data-sets	Fix: Download an older set of arxiv categories	4 months ago
 exercises	Add language detection exercise section	last week
 img	add learning curve plot	2 years ago
 presentations	Add files via upload	2 years ago
 transcripts	Update annif-install-codespaces.md	2 weeks ago
 .dockerignore	Move Dockerfile to master branch	4 years ago
 .gitignore	Ignore mauidata directory	5 years ago
 Dockerfile	Update Docker image to Annif 1.2	last week
 LICENSE.txt	Add CC BY 4.0 International license	5 years ago
 README.md	Own compute resources are not needed for Codespaces	last week

 **README**  CC-BY-4.0 license  

About






Instructions, exercises and example data sets for Annif hands-on tutorial

[machine-learning](#) [tutorial](#) [workshop](#) [text-classification](#) [glam](#) [code4lib](#) [multilabel-classification](#) [subject-indexing](#) [annif](#)

 Readme  CC-BY-4.0 license  Activity  Custom properties  36 stars  10 watching  9 forks [Report repository](#)

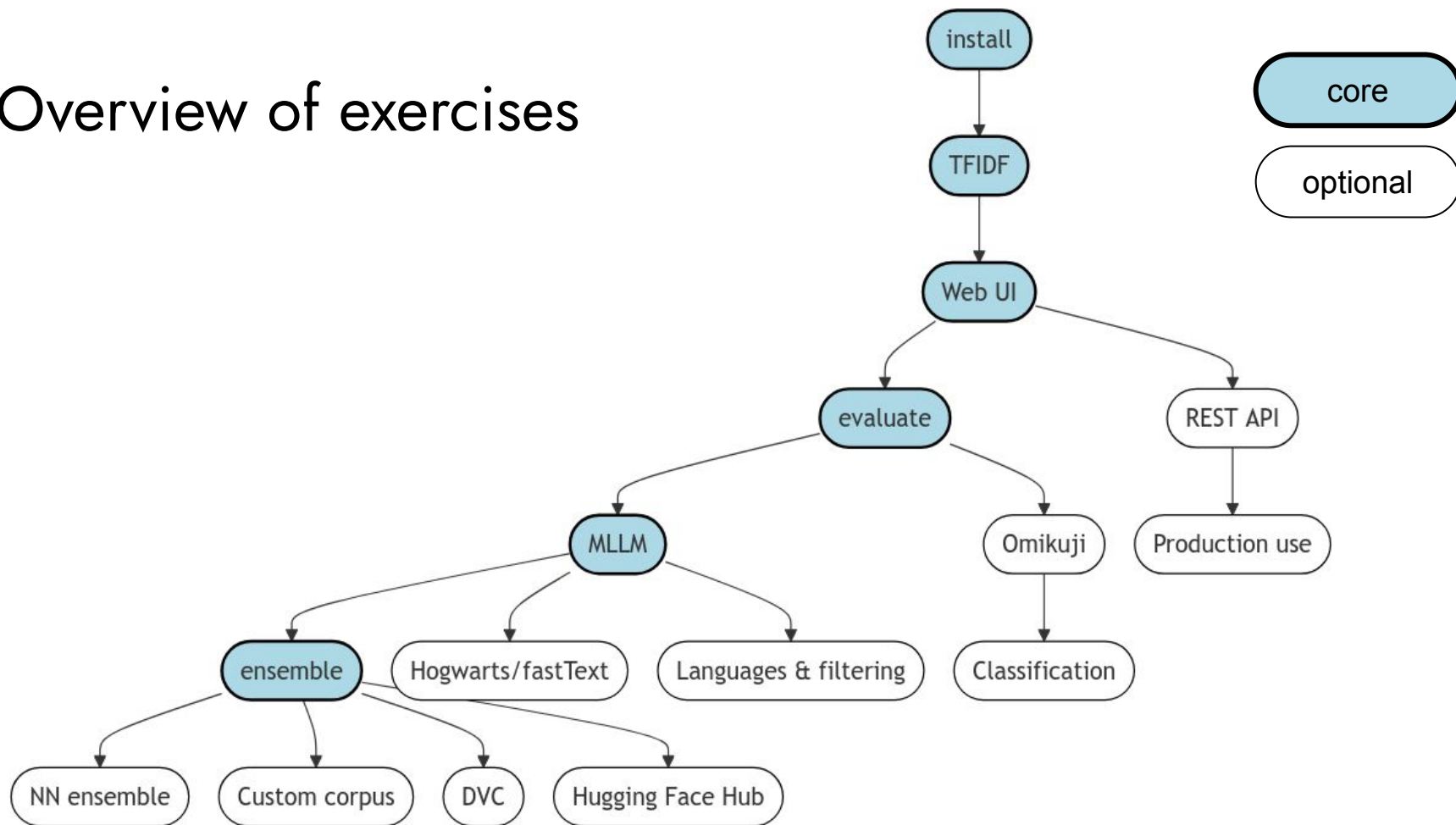
Contributors

5

Languages

Overview of exercises



Annif installation types

VirtualBox

Recommended for most people, as it's the easiest way of getting Annif running so you can work on the tutorial exercises.

Need to install

VirtualBox software - available for Windows, macOS and Linux.

Codespaces

If you can't install software on your own machine, you can use GitHub Codespaces to create a virtual machine and run Annif there.

You will need a free account on **GitHub**.

Docker

If you know Docker, this is a good way of getting Annif set up, with all the dependencies included in a pre-built container.

Need to install **Docker** software - available for Windows, macOS and Linux.

Local Linux

If you're an experienced Linux user and used to working with Python packages, a local install allows maximum flexibility.

Needs a recent version of **Python** with support for virtual environments.

Accessing Annif

Command line interface

- setup and administration
- training models
- testing and evaluating models
- bulk indexing of documents

Web user interface

- interactive testing of models

REST API

- integrating Annif services to other systems



A Short Introduction to the Linux Command Line

During the Tutorial Annif will mainly be controlled from Linux command line. This document provides an overview of the functionalities used. You can return to the examples in this document to better understand the commands used during the tutorial. You get access to the command line by opening the **Terminal** application.

Directories

Files in Linux are organized in a hierarchy of directories. You can navigate through the hierarchy from the command line. This is similar to clicking through folders in a graphical file explorer (i.e., Windows Explorer or Finder on Mac OS).

The Current Directory

Commands are usually executed relative to a directory. To show the current directory the `pwd` command can be used.

```
pwd
```

Contents of a Directory

The `ls` command, shorthand for *list*, lists the contents of the directory.

```
ls
```

Apply Annif on your own data!



Choose subject vocabulary



Prepare a corpus from
training data



Load the vocabulary and
train a model



Suggest subjects for new
documents