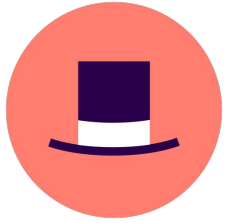


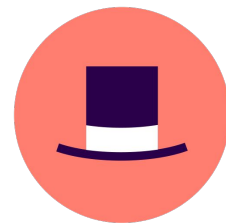
05. Maui



05. Set up and train a *Maui* project



Let's set up a *Maui* project. *Maui* is a great tool for lexical automated subject indexing, i.e. matching terms in a document text to terms in a controlled vocabulary.



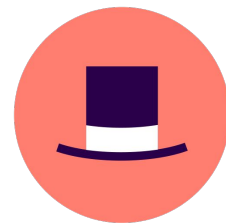
05. Set up and train a Maui project

In Annif, Maui can be used through Maui Server, a wrapper around the Maui codebase.

This means that Maui Server needs to be installed.

If you are using VirtualBox, the installation has been taken care of for you.

There is more information on Maui installation in previous videos and on the tutorial page.



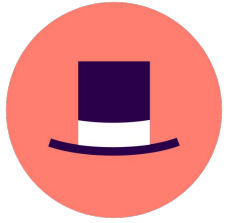
Install Maui: Docker

Open a new terminal window and run this:

```
docker run -v Maui-data-volume:/mauidata -p 8080:8080 --name mauiserver --rm -e  
MAUI_SERVER_DATA_DIR=/mauidata quay.io/natlibfi/mauiserver
```

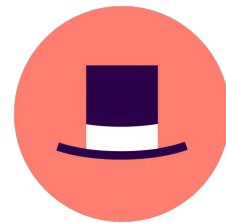
The above command will download the Docker image and start Maui. This might be the easiest option if you are already using Docker.

Install Maui: Linux



<https://github.com/NatLibFi/Annif/wiki/Backend%3A-Maui>

Linux install also covered in another tutorial video.



Test your setup

Try to access the URL <http://localhost:8080/mauiserver/> using a web browser or a tool such as curl.

You should get a (JSON) response like this:

annif-tutorial [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Mozilla Firefox [Terminal - annif@a... 21 Aug, 13:38

Mozilla Firefox

localhost:8080/mauiserver/

GitHub - NatLibFi/Annif-tutorial

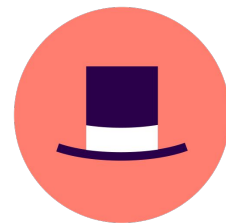
localhost:8080/mauiserver/

JSON Raw Data Headers

Save Copy Collapse All Expand All Filter JSON

```
title: "Maui Server"
data_dir: "/var/lib/mauidata"
default_lang: "en"
version: "1.3.2"
taggers: []
```

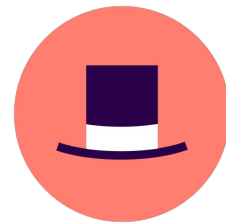
OIKEA CTRL



Step 1: Edit the projects.cfg -file

```
[yso-maui-en]  
name=YSO Maui project  
language=en  
backend=maui  
vocab=yso-en  
endpoint=http://localhost:8080/mauiserver/  
tagger=yso-maui-en
```

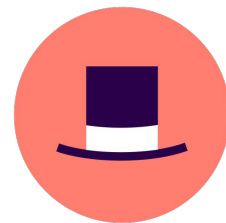
```
[stw-maui-en]  
name=STW Maui project  
language=en  
backend=maui  
vocab=stw-en  
endpoint=http://localhost:8080/mauiserver/  
tagger=stw-maui-en
```

Step 1: Edit the projects.cfg -file (for Docker)

```
[yso-maui-en]  
name=YSO Maui project  
language=en  
backend=maui  
vocab=yso-en  
endpoint=http://host.docker.internal:8080/mauiserver/  
tagger=yso-maui-en
```

```
[stw-maui-en]  
name=STW Maui project  
language=en  
backend=maui  
vocab=stw-en  
endpoint=http://host.docker.internal:8080/mauiserver/  
tagger=stw-maui-en
```



Step 2: Training

Maui requires a relatively small number (hundreds or at most a few thousand) of training documents.

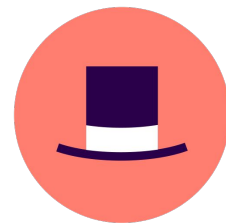
As usual, they should be as similar as possible in structure to the documents it will later be applied on. Note that we use fulltext documents in the validate folder for this:

```
annif train yso-maui-en data-sets/yso-nlf/docs/validate/
```

```
annif train stw-maui-en data-sets/stw-zbw/docs/validate/
```

```
(annif-venv) annif@annif-tutorial:~/Annif-tutorial$ annif train yso-maui-en data
-sets/yso-nlf/docs/validate/
Backend maui: Initializing Maui Server tagger 'yso-maui-en'
Backend maui: Uploading vocabulary
Backend maui: Creating train file
Backend maui: Uploading training documents
Backend maui: Waiting for training to be completed...
Backend maui: Training completed.
(annif-venv) annif@annif-tutorial:~/Annif-tutorial$
```

Step 3: Test w/ a sample text using annif suggest

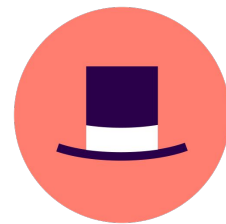


echo "frequently occurring or otherwise salient terms in the document are matched with terms in the vocabulary" | annif suggest yso-maui-en

echo "frequently occurring or otherwise salient terms in the document are matched with terms in the vocabulary" | annif suggest stw-maui-en

```
annif-venv) annif@annif-tutorial:~/Annif-tutorial$ echo "frequently occurring or otherwise salient terms in the document are matched with terms in the vocabulary" | annif suggest yso-maui-en
http://www.yso.fi/onto/yso/p2325>      documents      0.11600797840310018
http://www.yso.fi/onto/yso/p28134>      becoming more common      0.038067943456775326
http://www.yso.fi/onto/yso/p21428>      occurrence      0.033984234610338074
http://www.yso.fi/onto/yso/p21164>      combining      0.015163199541171962
http://www.yso.fi/onto/yso/p15046>      semesters      0.015163199541171962
http://www.yso.fi/onto/yso/p1346>      terminology      0.015163199541171962
annif-venv) annif@annif-tutorial:~/Annif-tutorial$
```

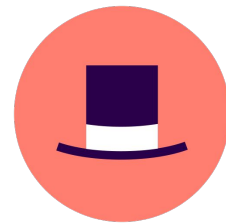
Step 4: Test w/ a document using annif suggest



```
annif suggest yso-maui-en <data-sets/yso-nlf/docs/test/2017-D-52518.txt
```

```
annif suggest stw-maui-en <data-sets/stw-zbw/docs/test/10008797547.txt
```

```
(annif-venv) annif@annif-tutorial:~/Annif-tutorial$ annif suggest yso-maui-en <data-sets/ys
/docs/test/2017-D-52518.txt
<http://www.yso.fi/onto/yso/p29466>      eusociality      0.3571695945647163
<http://www.yso.fi/onto/yso/p29456>      kin selection    0.29257174044188305
<http://www.yso.fi/onto/yso/p9667>       queens          0.23634235666851292
<http://www.yso.fi/onto/yso/p21526>      subordinates     0.21903030303030305
<http://www.yso.fi/onto/yso/p3510>       colony          0.19348521381137004
<http://www.yso.fi/onto/yso/p13476>      game theory     0.17355380982325358
<http://www.yso.fi/onto/yso/p1589>       nests           0.15918979658491836
<http://www.yso.fi/onto/yso/p24562>      resource allocation 0.1447861926563353
<http://www.yso.fi/onto/yso/p27928>      focalization    0.14330117620486726
<http://www.yso.fi/onto/yso/p3837>       colonialism     0.13967569000184624
(annif-venv) annif@annif-tutorial:~/Annif-tutorial$
```



Step 5: Test w/ a corpus using annif eval

```
annif eval yso-maui-en data-sets/yso-nlf/docs/test/
```

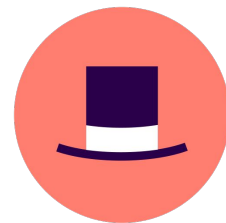
```
annif eval stw-maui-en data-sets/stw-zbw/docs/test/
```



```
(annif-venv) annif@annif-tutorial:~/Annif-tutorial$ annif eval yso-maui-en data-sets/yso-nlf/docs/tes
Precision (doc avg):          0.235
Recall (doc avg):             0.37595645150056917
F1 score (doc avg):           0.2747838286073501
Precision (subj avg):         0.012086141346797344
Recall (subj avg):            0.01364758530379172
F1 score (subj avg):          0.012122255258984428
Precision (weighted subj avg): 0.31755427517115237
Recall (weighted subj avg):    0.3432327166504382
F1 score (weighted subj avg): 0.3056918510341264
Precision (microavg):         0.235
Recall (microavg):            0.3432327166504382
F1 score (microavg):          0.2789869410368025
F1@5:                          0.29324072078426183
NDCG:                          0.39816760722194294
NDCG@5:                        0.41117058089773106
NDCG@10:                       0.4075691916037827
```

```
Precision@1:          0.5866666666666667
Precision@3:          0.4144444444444445
Precision@5:          0.338
LRAP:                 0.25123405101702445
True positives:       705
False positives:      2295
False negatives:      1349
Documents evaluated:   300
(annif-venv) annif@annif-tutorial:~/Annif-tutorial$
```

Step 6: All done!



How did your tests go? Note that the ensemble backend is available in the Annif UI, and you can test it as well.