

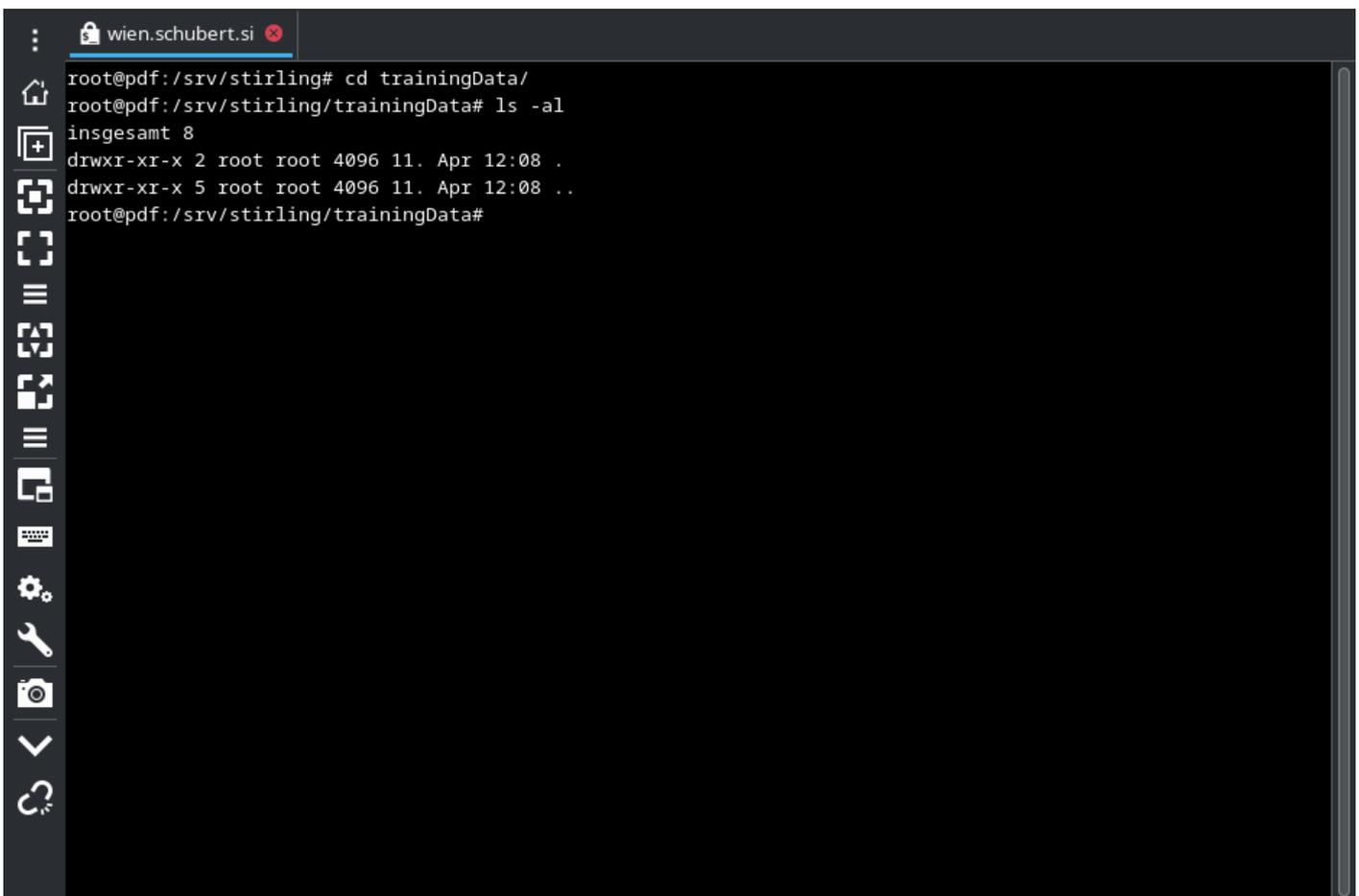
Sprachpakete hinzufügen

In der Grundinstallation verfügt Stirling-PDF nur über das englischsprachige Paket für die Texterkennung.

Um bspw. deutschsprachige Texte verarbeiten zu können, bedarf es somit weiterer Sprachpakete.

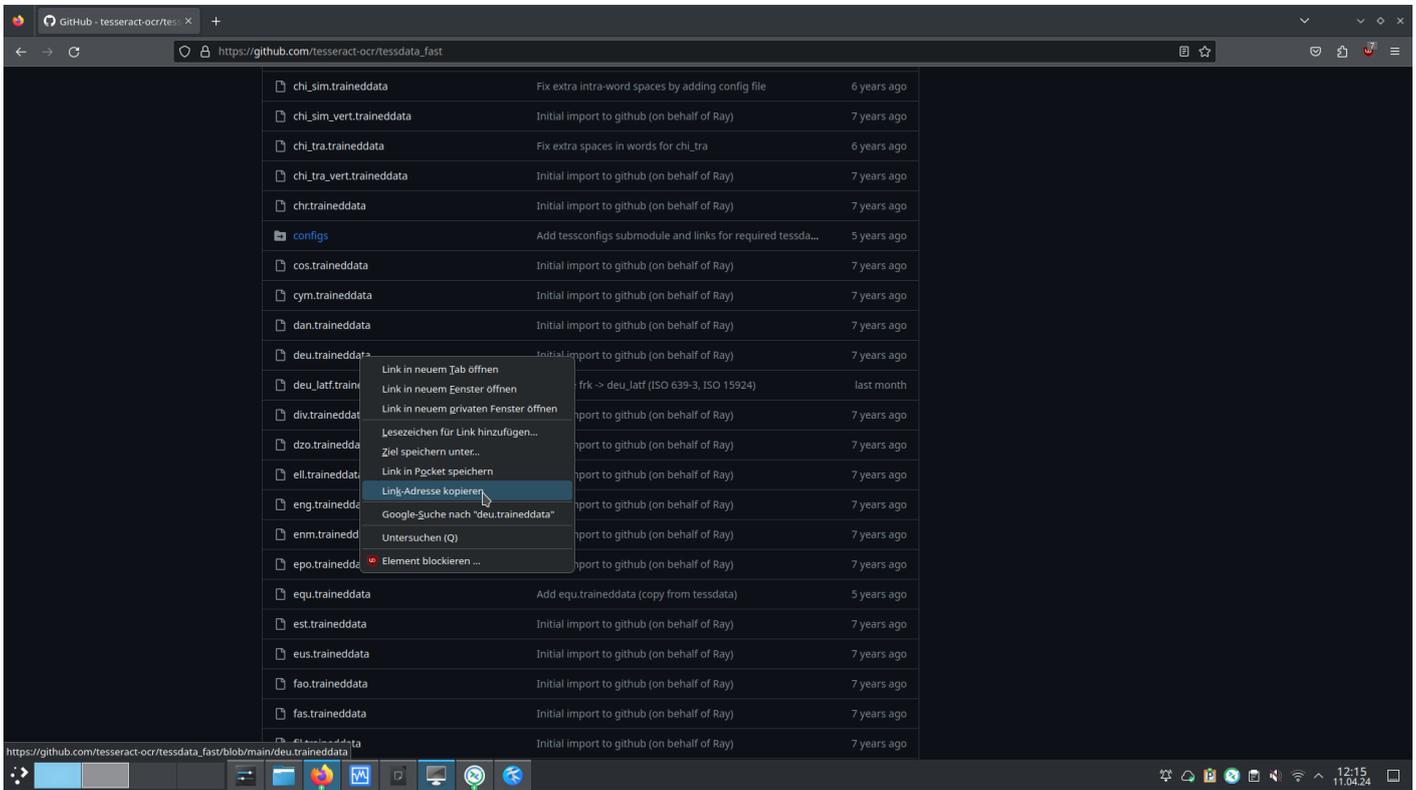
Die optionalen Sprachpakete befinden sich unterhalb des Installationsverzeichnis (hier `/srv/stirling`) im Verzeichnis `trainingData`.

Daher wird zunächst in dieses Verzeichnis gewechselt, das nach der Installation noch leer ist:

A terminal window with a dark background and light text. The window title is 'wien.schubert.si'. The terminal shows the following commands and output:

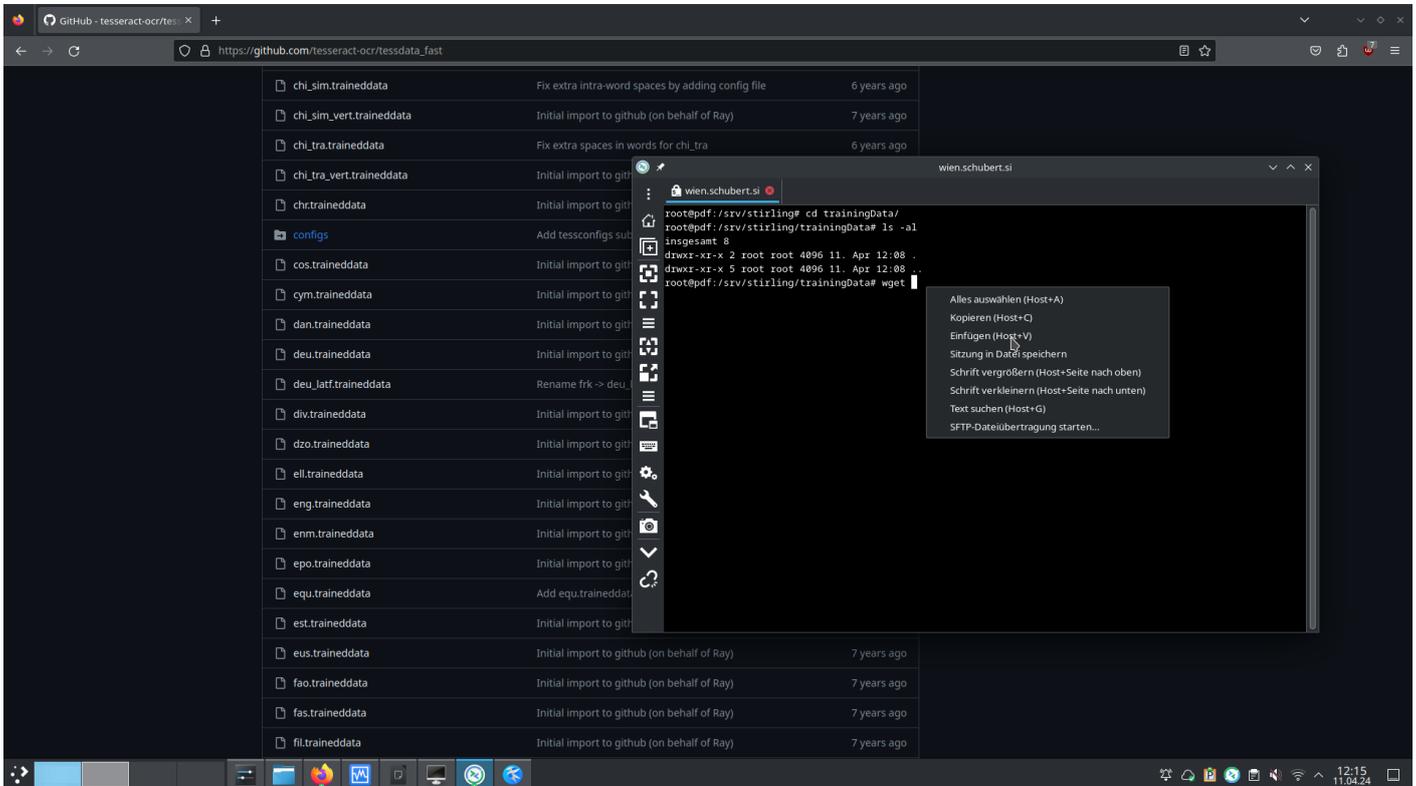
```
root@pdf:/srv/stirling# cd trainingData/
root@pdf:/srv/stirling/trainingData# ls -al
insgesamt 8
drwxr-xr-x 2 root root 4096 11. Apr 12:08 .
drwxr-xr-x 5 root root 4096 11. Apr 12:08 ..
root@pdf:/srv/stirling/trainingData#
```

Auf der [Übersichtsseite aller verfügbaren Sprachpakete](#) wird die URL des zu installierenden Paketes (in diesem Beispiel `deutsch`) ermittelt und kopiert:



Im Verzeichnis der Sprachpakete kann die gewünschte Datei dann mittels `wget URL` empfangen werden.

Für das Beispiel lautet der gesamte Befehl somit `wget https://github.com/tesseract-ocr/tessdata_fast/blob/main/deu.traineddata`.



Nach dem Download enthält das Verzeichnis *trainingData* die entsprechende Sprachdatei:

```
wien.schubert.si x
root@pdf:/srv/stirling/trainingData# ls -al
insgesamt 164
drwxr-xr-x 2 root root 4096 11. Apr 12:23 .
drwxr-xr-x 5 root root 4096 11. Apr 12:08 ..
-rw-r--r-- 1 root root 156627 11. Apr 12:23 deu.traineddata
root@pdf:/srv/stirling/trainingData#
```

Obwohl die Datei jetzt vorhanden ist, weiß die Anwendung (Stirling-PDF) noch nichts von ihr. Hierfür muss abschließend der Container neu gestartet werden.

Im ersten Schritt wird die laufende Anwendung per `docker-compose down` heruntergefahren und im zweiten Schritt durch `docker-compose up -d` wieder gestartet.

```
wien.schubert.si x
root@pdf:/srv/stirling/trainingData# docker-compose down
Removing stirling_stirling-pdf_1 ... done
Removing network stirling_default
root@pdf:/srv/stirling/trainingData# docker-compose up -d
Creating network "stirling_default" with the default driver
Creating stirling_stirling-pdf_1 ... done
root@pdf:/srv/stirling/trainingData#
```

Das zusätzliche Sprachpaket steht Stirling-PDF ab sofort zur Verfügung.

Revision #2

Created 11 April 2024 12:41:53 by Rene Schubert

Updated 11 April 2024 13:45:59 by Rene Schubert