

Installation

Diese Schritt-für-Schritt-Anleitung beschreibt alle Maßnahmen, die bei einem frisch installierten Debian Linux (hier: Version 12.5 Bookworm) nötig sind, um die Anwendung (extern) nutzen zu können.

- [Konfiguration des Docker-Containers](#)
- [Installation von Docker und Docker-Compose](#)
- [Docker-Container erstellen](#)
- [Sprachpakete hinzufügen](#)

Konfiguration des Docker-Containers

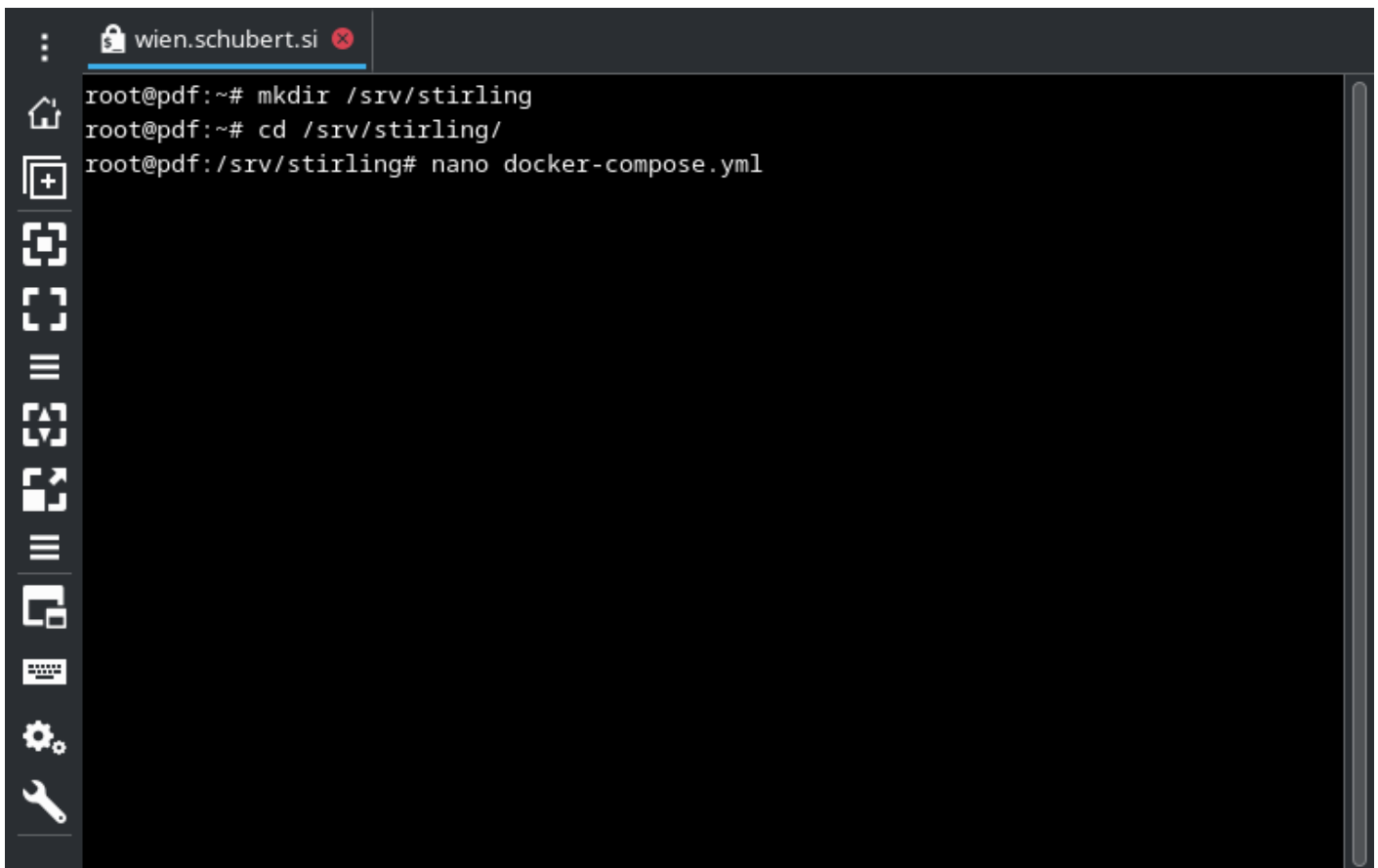
Wie bei [Docker-Compose](#) üblich wird der Container mit Hilfe einer speziellen Konfigurationsdatei (*docker-compose.yml*) spezifiziert.

Neben den Dateien des Docker-Containers sind zudem noch weitere Daten, bspw. in Form von Trainingsdaten für die Texterkennung, zu erwarten.

Um die Organisation der Dateien auf dem Server übersichtlich zu halten, empfiehlt es sich grundsätzlich für jeden (Server-)Dienst ein separates Verzeichnis zu nutzen.

In diesem Beispiel wird daher zunächst mit dem Befehl `mkdir /srv/stirling` ein neues Verzeichnis für Stirling-PDF erstellt und mittels `cd /srv/stirling/` dorthin gewechselt.

Dort wird dann mit Hilfe eines Text-Editors die Konfigurationsdatei angelegt. Bei der Verwendung von *nano* lautet der Befehl somit `nano docker-compose.yml`:



```
wien.schubert.si x
root@pdf:~# mkdir /srv/stirling
root@pdf:~# cd /srv/stirling/
root@pdf:/srv/stirling# nano docker-compose.yml
```

Die Konfigurationsdatei wird mit Daten nach diesem Schema erstellt:

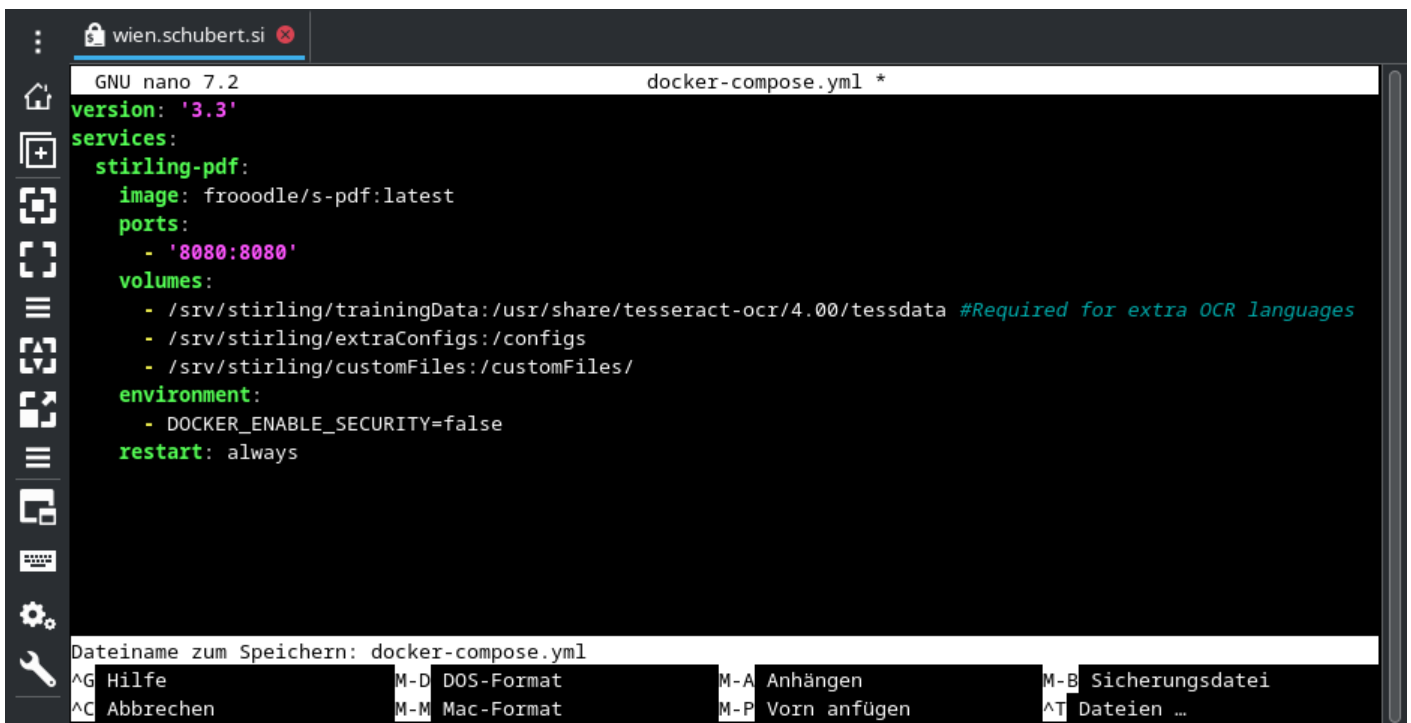
```
version: '3.3'
services:
  stirring-pdf:
    image: frooodle/s-pdf:latest
    ports:
      - '8080:8080'
    volumes:
      - /srv/stirling/trainingData:/usr/share/tesseract-ocr/4.00/tessdata #Required for extra OCR languages
      - /srv/stirling/extraConfigs:/configs
      - /srv/stirling/customFiles:/customFiles/
    environment:
      - DOCKER_ENABLE_SECURITY=false
    restart: always
```

Unter *ports* wird definiert, über welchen Port (links anzugeben) die Anwendung, die im Container unter dem rechts angegebenen Port läuft, später von außen zu erreichen ist. Der Eintrag `:8080` darf nicht verändert werden.

Soll Stirling-PDF später unter dem üblichen HTTP-Port 80 erreichbar sein, müsste der Eintrag auf `'80:8080'` geändert werden.

Bei *volumes* muss der Pfad angepasst werden, wenn nicht wie in diesem Beispiel `/srv/stirling` verwendet wird.

Mit der Tastenkombination `STRG` und `X` kann der Editor *nano* verlassen und die Datei nach Rückfrage gespeichert werden:



```
GNU nano 7.2 docker-compose.yml *
version: '3.3'
services:
  stirring-pdf:
    image: frooodle/s-pdf:latest
    ports:
      - '8080:8080'
    volumes:
      - /srv/stirling/trainingData:/usr/share/tesseract-ocr/4.00/tessdata #Required for extra OCR languages
      - /srv/stirling/extraConfigs:/configs
      - /srv/stirling/customFiles:/customFiles/
    environment:
      - DOCKER_ENABLE_SECURITY=false
    restart: always

Dateiname zum Speichern: docker-compose.yml
^G Hilfe      M-D DOS-Format  M-A Anhängen   M-B Sicherungsdatei
^C Abbrechen  M-M Mac-Format  M-P Vorn anfügen ^T Dateien ...
```

Die Definition des Docker-Containers ist damit abgeschlossen.

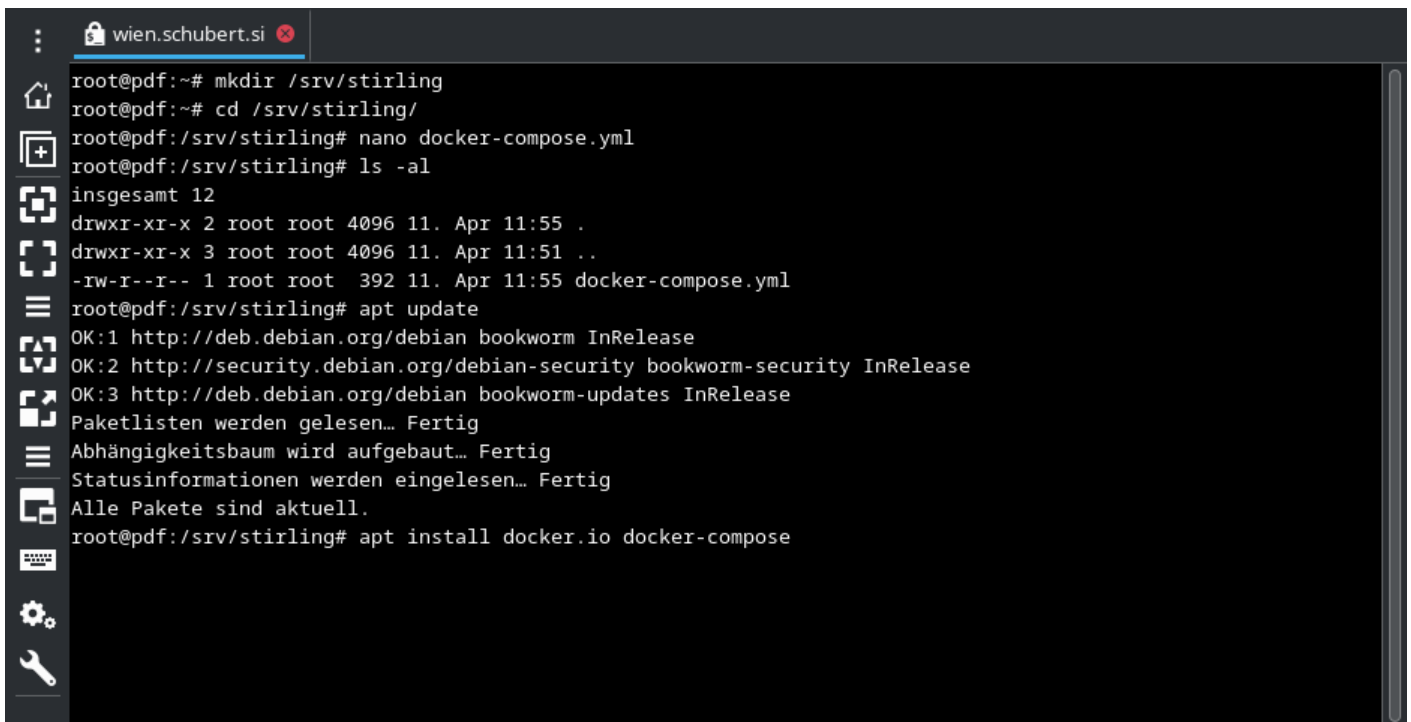
Installation von Docker und Docker-Compose

Auf einem frisch aufgesetzten Linux-Betriebssystem (wie hier Debian 12.5 Bookworm) ist in der Regel außer SSH kein weiterer Dienst - und somit auch kein [Docker](#) - verfügbar.

Daher gilt es zunächst Docker ("Betreiber des Containers") und Docker-Compose ("Zusammenbauer des Containers") in Betrieb zu nehmen.

Dafür werden als erstes die Paketquellen mittels `apt update` aktualisiert und ggf. Updates per `apt upgrade` (in diesem Beispiel nicht erforderlich, weil alle Pakete aktuell sind) eingespielt.

Im Anschluss daran werden Docker und Docker-Compose durch den Befehl `apt install docker.io docker-compose` gemeinsam installiert:

A terminal window titled 'wien.schubert.si' showing a series of commands and their outputs. The user creates a directory, navigates to it, creates a file, and lists its contents. Then, they run 'apt update' which shows that all packages are up to date. Finally, they run 'apt install docker.io docker-compose' to install the required software.

```
root@pdf:~# mkdir /srv/stirling
root@pdf:~# cd /srv/stirling/
root@pdf:/srv/stirling# nano docker-compose.yml
root@pdf:/srv/stirling# ls -al
insgesamt 12
drwxr-xr-x 2 root root 4096 11. Apr 11:55 .
drwxr-xr-x 3 root root 4096 11. Apr 11:51 ..
-rw-r--r-- 1 root root 392 11. Apr 11:55 docker-compose.yml
root@pdf:/srv/stirling# apt update
OK:1 http://deb.debian.org/debian bookworm InRelease
OK:2 http://security.debian.org/debian-security bookworm-security InRelease
OK:3 http://deb.debian.org/debian bookworm-updates InRelease
Paketlisten werden gelesen... Fertig
Abhängigkeitsbaum wird aufgebaut... Fertig
Statusinformationen werden eingelesen... Fertig
Alle Pakete sind aktuell.
root@pdf:/srv/stirling# apt install docker.io docker-compose
```

Nach der Bestätigung mit der Taste `J` (oder `Y`, wenn keine dt. Lokalisation vorliegt) ...

```
wien.schubert.si x
root@pdf:/srv/stirling# apt install docker.io docker-compose
Paketlisten werden gelesen... Fertig
Abhängigkeitsbaum wird aufgebaut... Fertig
Statusinformationen werden eingelesen... Fertig
Die folgenden zusätzlichen Pakete werden installiert:
 binutils binutils-common binutils-x86-64-linux-gnu cgroupfs-mount containerd criu git git-man libbinutils
 libctf-nobfd0 libctf0 liberror-perl libgprofng0 libint1-perl libint1-xs-perl libmodule-find-perl
 libmodule-scandeps-perl libnet1 libnl-3-200 libproc-processtable-perl libprotobuf32
 libsort-naturally-perl libterm-readkey-perl libyaml-0-2 needrestart patch python3-attr python3-distro
 python3-distutils python3-docker python3-dockerpty python3-docopt python3-dotenv python3-json-pointer
 python3-jsonschema python3-lib2to3 python3-protobuf python3-pyrsistent python3-rfc3987 python3-texttable
 python3-uritemplate python3-webcolors python3-websocket python3-yaml runc sgml-base tini
Vorgeschlagene Pakete:
 binutils-doc containerd-networking-plugins docker-doc aufs-tools btrfs-progs debootstrap rinse rootlesskit
 xfsprogs zfs-fuse | zfsutils-linux git-daemon-run | git-daemon-sysvinit git-doc git-email git-gui gitk
 gitweb git-cvs git-mediawiki git-svn needrestart-session | libnotify-bin ed diffutils-doc python-attr-doc
 python-jsonschema-doc sgml-base-doc
Die folgenden NEUEN Pakete werden installiert:
 binutils binutils-common binutils-x86-64-linux-gnu cgroupfs-mount containerd criu docker-compose
 docker.io git git-man libbinutils libctf-nobfd0 libctf0 liberror-perl libgprofng0 libint1-perl
 libint1-xs-perl libmodule-find-perl libmodule-scandeps-perl libnet1 libnl-3-200 libproc-processtable-perl
 libprotobuf32 libsort-naturally-perl libterm-readkey-perl libyaml-0-2 needrestart patch python3-attr
 python3-distro python3-distutils python3-docker python3-dockerpty python3-docopt python3-dotenv
 python3-json-pointer python3-jsonschema python3-lib2to3 python3-protobuf python3-pyrsistent
 python3-rfc3987 python3-texttable python3-uritemplate python3-webcolors python3-websocket python3-yaml
 runc sgml-base tini
0 aktualisiert, 49 neu installiert, 0 zu entfernen und 0 nicht aktualisiert.
Es müssen 84,8 MB an Archiven heruntergeladen werden.
Nach dieser Operation werden 361 MB Plattenplatz zusätzlich benutzt.
Möchten Sie fortfahren? [J/n] j
```

... werden die Pakete heruntergeladen, entpackt ...

```
wien.schubert.si x
Holen:32 http://deb.debian.org/debian bookworm/main amd64 python3-yaml amd64 6.0-3+b2 [119 kB]
Holen:33 http://deb.debian.org/debian bookworm/main amd64 docker-compose all 1.29.2-3 [123 kB]
Holen:34 http://deb.debian.org/debian bookworm/main amd64 liberror-perl all 0.17029-2 [29,0 kB]
Holen:35 http://deb.debian.org/debian bookworm/main amd64 git-man all 1:2.39.2-1.1 [2.049 kB]
Holen:36 http://deb.debian.org/debian bookworm/main amd64 git amd64 1:2.39.2-1.1 [7.171 kB]
Holen:37 http://deb.debian.org/debian bookworm/main amd64 libint1-perl all 1.33-1 [720 kB]
Holen:38 http://deb.debian.org/debian bookworm/main amd64 libint1-xs-perl amd64 1.33-1 [15,6 kB]
Holen:39 http://deb.debian.org/debian bookworm/main amd64 libmodule-find-perl all 0.16-2 [10,6 kB]
Holen:40 http://deb.debian.org/debian bookworm/main amd64 libmodule-scandeps-perl all 1.31-2 [41,7 kB]
Holen:41 http://deb.debian.org/debian bookworm/main amd64 libproc-processtable-perl amd64 0.634-1+b2 [43,1 kB]
]
Holen:42 http://deb.debian.org/debian bookworm/main amd64 libsort-naturally-perl all 1.03-4 [13,1 kB]
Holen:43 http://deb.debian.org/debian bookworm/main amd64 libterm-readkey-perl amd64 2.38-2+b1 [24,5 kB]
Holen:44 http://deb.debian.org/debian bookworm/main amd64 needrestart all 3.6-4+deb12u1 [59,8 kB]
Holen:45 http://deb.debian.org/debian bookworm/main amd64 patch amd64 2.7.6-7 [128 kB]
Holen:46 http://deb.debian.org/debian bookworm/main amd64 python3-json-pointer all 2.3-2 [15,1 kB]
Holen:47 http://deb.debian.org/debian bookworm/main amd64 python3-rfc3987 all 1.3.8-2 [8.816 B]
Holen:48 http://deb.debian.org/debian bookworm/main amd64 python3-uritemplate all 4.1.1-2 [10,9 kB]
Holen:49 http://deb.debian.org/debian bookworm/main amd64 python3-webcolors all 1.11.1-1 [12,7 kB]
Es wurden 84,8 MB in 2 s geholt (48,6 MB/s).
Extrahiere Vorlagen aus Paketen: 100%
Vormals nicht ausgewähltes Paket runc wird gewählt.
(Lese Datenbank ... 30057 Dateien und Verzeichnisse sind derzeit installiert.)
Vorbereitung zum Entpacken von .../00-runc_1.1.5+ds1-1+deb12u1_amd64.deb ...
Entpacken von runc (1.1.5+ds1-1+deb12u1) ...
Vormals nicht ausgewähltes Paket containerd wird gewählt.
Vorbereitung zum Entpacken von .../01-containerd_1.6.20~ds1-1+b1_amd64.deb ...
Entpacken von containerd (1.6.20~ds1-1+b1) ...

Fortschritt: [ 1%] [.....]
```

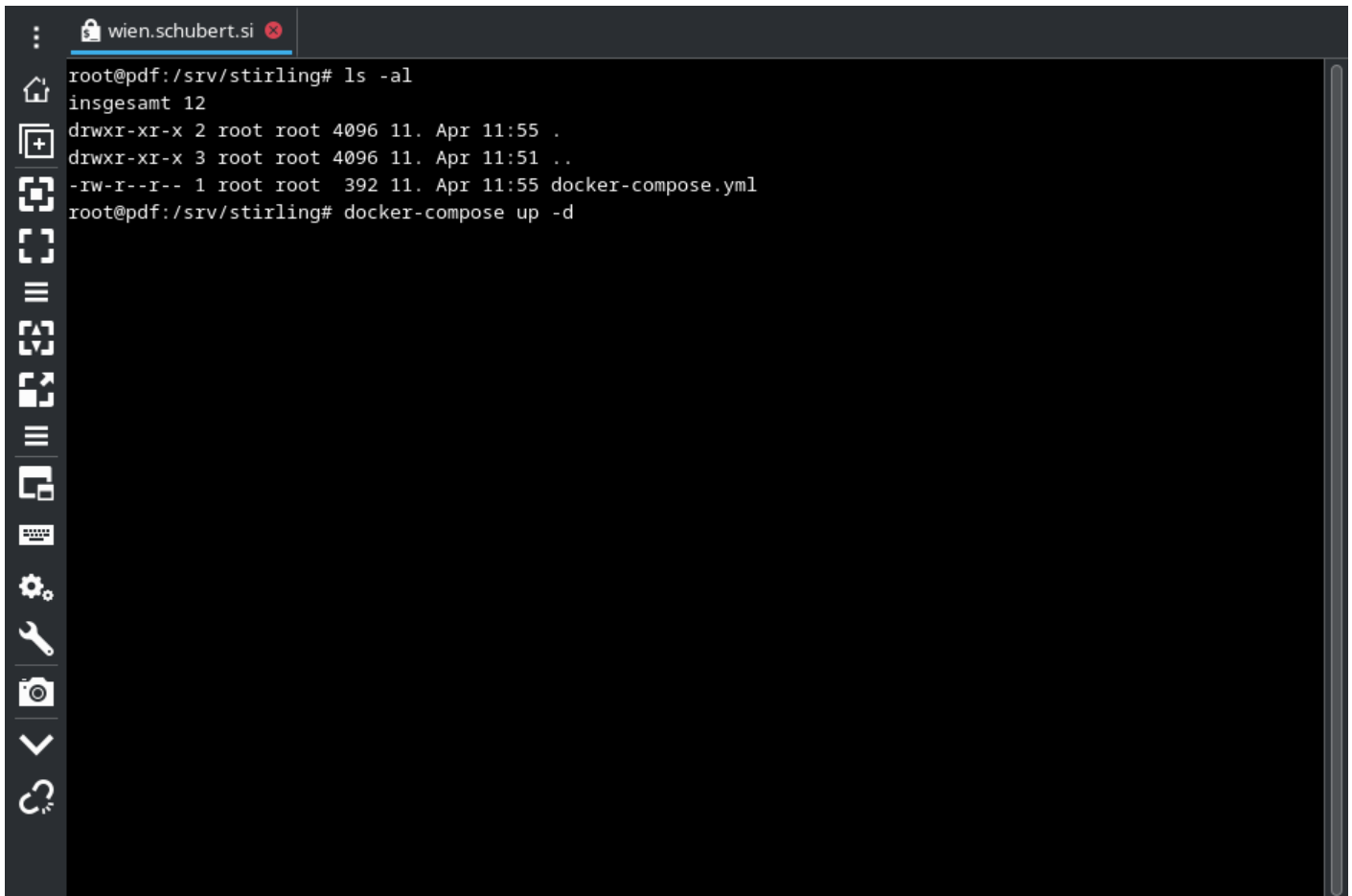
... und eingerichtet:

```
wien.schubert.si x
rd.service.
python3-lib2to3 (3.11.2-3) wird eingerichtet ...
libsort-naturally-perl (1.03-4) wird eingerichtet ...
python3-websocket (1.2.3-1) wird eingerichtet ...
libmodule-scandeps-perl (1.31-2) wird eingerichtet ...
python3-dockerpty (0.4.1-4) wird eingerichtet ...
libc64 (2.40-2) wird eingerichtet ...
python3-distutils (3.11.2-3) wird eingerichtet ...
python3-docker (5.0.3-1) wird eingerichtet ...
docker.io (20.10.24+dfsg1-1+b3) wird eingerichtet ...
Lege Gruppe »docker« (GID 109) an ...
Fertig.
Created symlink /etc/systemd/system/multi-user.target.wants/docker.service → /lib/systemd/system/docker.service.
Created symlink /etc/systemd/system/sockets.target.wants/docker.socket → /lib/systemd/system/docker.socket.
libnet1:amd64 (1.1.6+dfsg-3.2) wird eingerichtet ...
libintl-xs-perl (1.33-1) wird eingerichtet ...
libprofng0:amd64 (2.40-2) wird eingerichtet ...
python3-jjsonschema (4.10.3-1) wird eingerichtet ...
git (1:2.39.2-1.1) wird eingerichtet ...
criu (3.17.1-2) wird eingerichtet ...
binutils-x86-64-linux-gnu (2.40-2) wird eingerichtet ...
docker-compose (1.29.2-3) wird eingerichtet ...
binutils (2.40-2) wird eingerichtet ...
needrestart (3.6-4+deb12u1) wird eingerichtet ...
Trigger für man-db (2.11.2-2) werden verarbeitet ...
Trigger für libc-bin (2.36-9+deb12u4) werden verarbeitet ...
root@pdf:/srv/stirling#
```

Docker und Docker-Compose sind jetzt betriebsbereit.

Docker-Container erstellen

In dem Verzeichnis, in dem auch die *docker-compose.yml* liegt, wird der Befehl `docker-compose up -d` abgesetzt:

A terminal window with a dark background and light text. The title bar shows 'wien.schubert.si' with a red close button. The terminal content shows a user at the prompt 'root@pdf:/srv/stirling#' running 'ls -al'. The output lists the current directory, its parent, and a file named 'docker-compose.yml' with permissions '-rw-r--r--'. The user then runs 'docker-compose up -d'.

```
root@pdf:/srv/stirling# ls -al
insgesamt 12
drwxr-xr-x 2 root root 4096 11. Apr 11:55 .
drwxr-xr-x 3 root root 4096 11. Apr 11:51 ..
-rw-r--r-- 1 root root  392 11. Apr 11:55 docker-compose.yml
root@pdf:/srv/stirling# docker-compose up -d
```

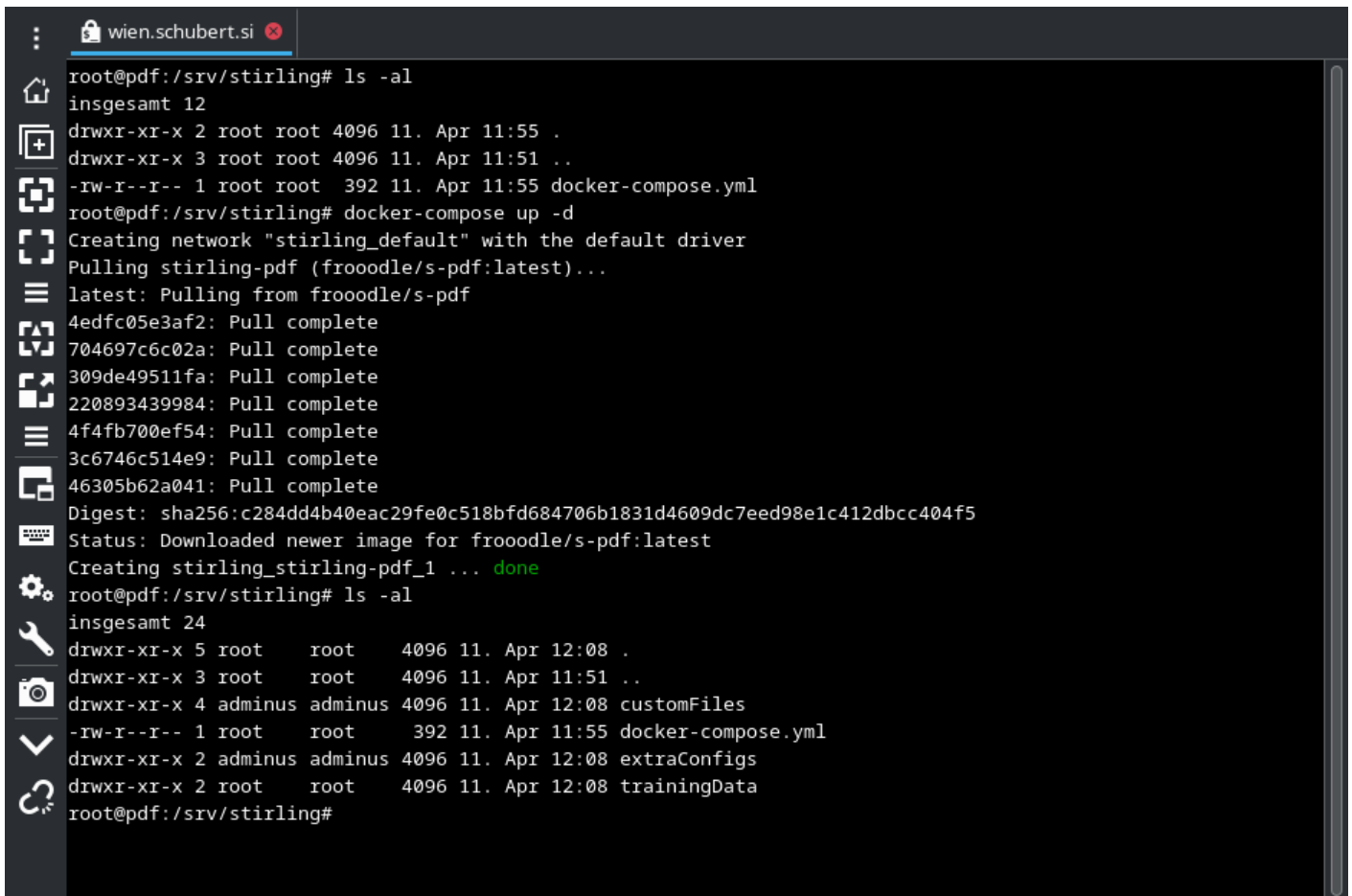
Daraufhin werden die benötigten Daten angefordert ...


```
wien.schubert.si x
root@pdf:/srv/stirling# ls -al
insgesamt 12
drwxr-xr-x 2 root root 4096 11. Apr 11:55 .
drwxr-xr-x 3 root root 4096 11. Apr 11:51 ..
-rw-r--r-- 1 root root 392 11. Apr 11:55 docker-compose.yml
root@pdf:/srv/stirling# docker-compose up -d
Creating network "stirling_default" with the default driver
Pulling stirling-pdf (frooodle/s-pdf:latest)...
latest: Pulling from frooodle/s-pdf
4edfc05e3af2: Pulling fs layer
704697c6c02a: Download complete
309de49511fa: Download complete
220893439984: Download complete
4f4fb700ef54: Download complete
3c6746c514e9: Download complete
46305b62a041: Downloading [=====>] 113.9MB/667.7MB
```

... ausgepackt ...

```
wien.schubert.si x
root@pdf:/srv/stirling# ls -al
insgesamt 12
drwxr-xr-x 2 root root 4096 11. Apr 11:55 .
drwxr-xr-x 3 root root 4096 11. Apr 11:51 ..
-rw-r--r-- 1 root root 392 11. Apr 11:55 docker-compose.yml
root@pdf:/srv/stirling# docker-compose up -d
Creating network "stirling_default" with the default driver
Pulling stirling-pdf (frooodle/s-pdf:latest)...
latest: Pulling from frooodle/s-pdf
4edfc05e3af2: Pull complete
704697c6c02a: Pull complete
309de49511fa: Pull complete
220893439984: Pull complete
4f4fb700ef54: Pull complete
3c6746c514e9: Pull complete
46305b62a041: Extracting [=====>] 349.3MB/667.7MB
```

... sowie der Container erstellt und in Betrieb genommen (Meldung **done**):

A terminal window with a dark background and light text. The window title is 'wien.schubert.si'. The user is at the prompt 'root@pdf:/srv/stirling#'. They run 'ls -al' showing 12 files. Then they run 'docker-compose up -d'. The output shows the creation of a network and pulling the 'stirling-pdf' image from 'frooodle/s-pdf:latest'. It lists several layers being pulled and their completion. A digest is shown. The status is 'Downloaded newer image for frooodle/s-pdf:latest'. Finally, 'Creating stirling_stirling-pdf_1 ... done' is shown. The user runs 'ls -al' again, showing 24 files, including 'customFiles', 'extraConfigs', and 'trainingData'.

```
root@pdf:/srv/stirling# ls -al
insgesamt 12
drwxr-xr-x 2 root root 4096 11. Apr 11:55 .
drwxr-xr-x 3 root root 4096 11. Apr 11:51 ..
-rw-r--r-- 1 root root 392 11. Apr 11:55 docker-compose.yml
root@pdf:/srv/stirling# docker-compose up -d
Creating network "stirling_default" with the default driver
Pulling stirling-pdf (frooodle/s-pdf:latest)...
latest: Pulling from frooodle/s-pdf
4edfc05e3af2: Pull complete
704697c6c02a: Pull complete
309de49511fa: Pull complete
220893439984: Pull complete
4f4fb700ef54: Pull complete
3c6746c514e9: Pull complete
46305b62a041: Pull complete
Digest: sha256:c284dd4b40eac29fe0c518bfd684706b1831d4609dc7eed98e1c412dbcc404f5
Status: Downloaded newer image for frooodle/s-pdf:latest
Creating stirling_stirling-pdf_1 ... done
root@pdf:/srv/stirling# ls -al
insgesamt 24
drwxr-xr-x 5 root root 4096 11. Apr 12:08 .
drwxr-xr-x 3 root root 4096 11. Apr 11:51 ..
drwxr-xr-x 4 adminus adminus 4096 11. Apr 12:08 customFiles
-rw-r--r-- 1 root root 392 11. Apr 11:55 docker-compose.yml
drwxr-xr-x 2 adminus adminus 4096 11. Apr 12:08 extraConfigs
drwxr-xr-x 2 root root 4096 11. Apr 12:08 trainingData
root@pdf:/srv/stirling#
```

Ab jetzt ist Stirling-PDF verfügbar.
(In diesem Beispiel via <http://IP-Adresse:8080>)

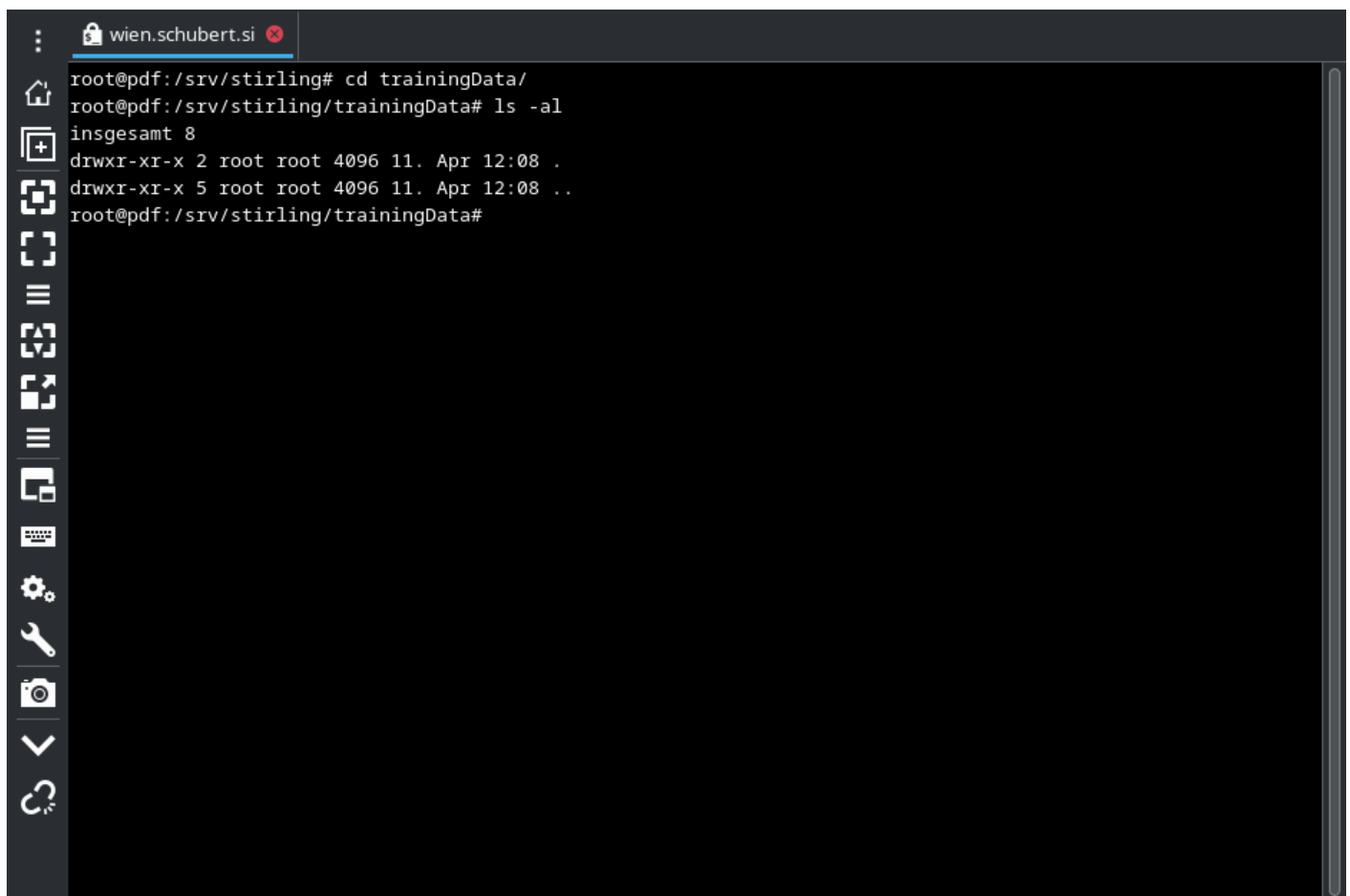
Sprachpakete hinzufügen

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.

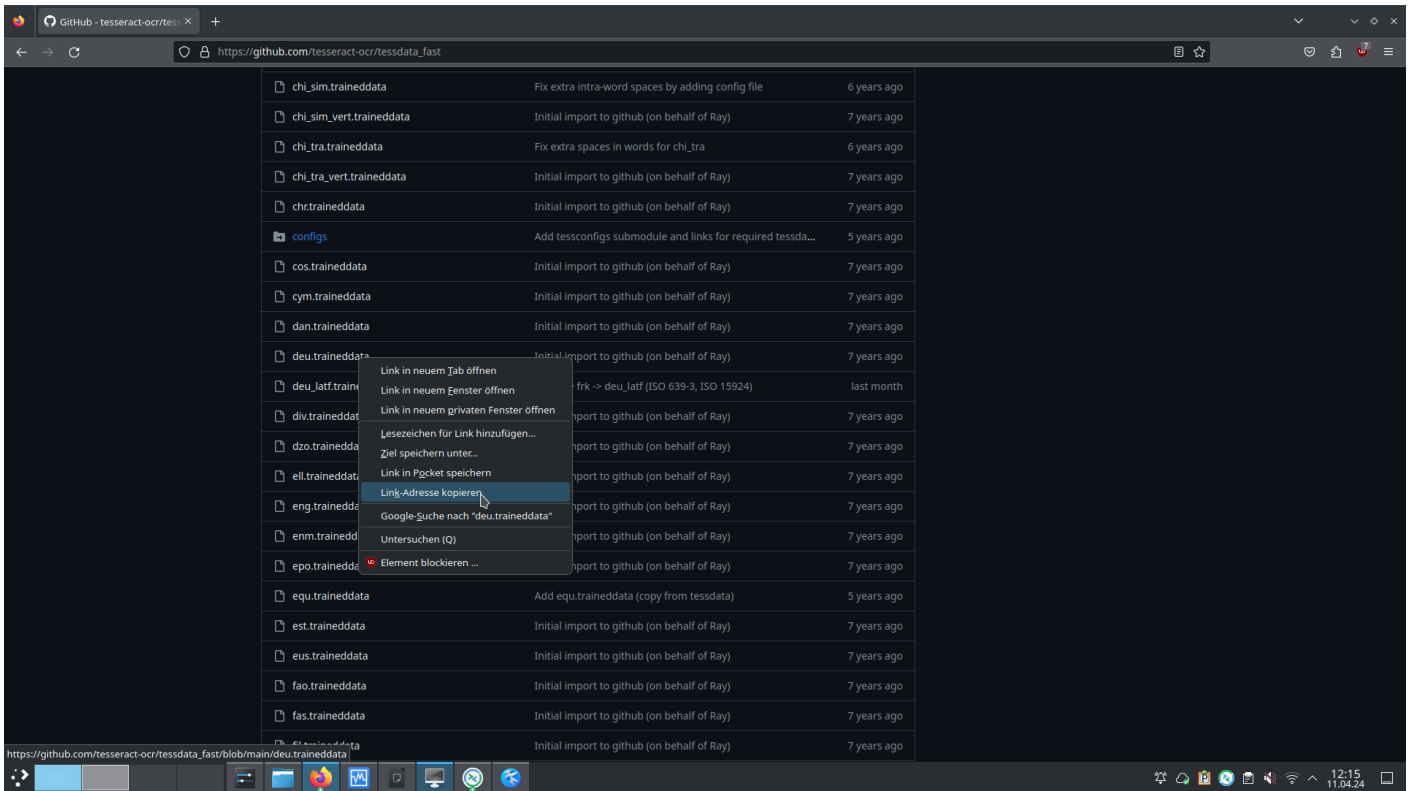
Die optionalen Sprachpakete befinden sich unterhalb des Installationsverzeichnis (hier */srv/stirling*) im Verzeichnis *trainingData*.

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 title bar shows 'wien.schubert.si' with a red close button. The terminal content shows a user at the root of a PDF server navigating to the 'trainingData' directory and listing its contents. The output shows two files: a file named '.' and a directory named '..', both with permissions 'drwxr-xr-x', owned by 'root', size '4096', and timestamp '11. Apr 12:08'.

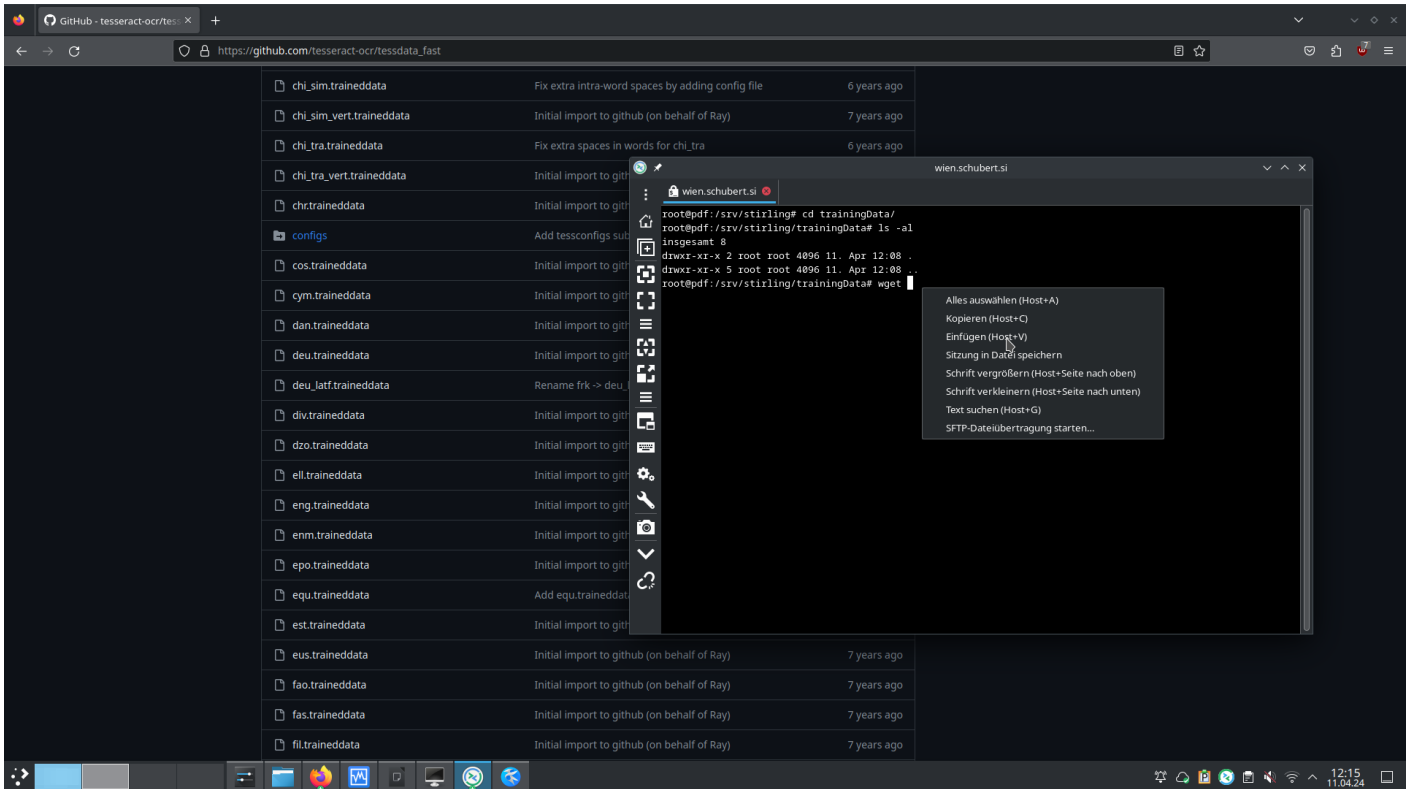
```
root@pdf:/srv/stirling# cd trainingData/
root@pdf:/srv/stirling/trainingData# ls -al
insgesamt 8
drwxr-xr-x 2 root root 4096 11. Apr 12:08 .
drwxr-xr-x 5 root root 4096 11. Apr 12:08 ..
root@pdf:/srv/stirling/trainingData#
```

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.