

# Requirements

Linux OS : Ubuntu 14.04 server eg. ubuntu-14.04.3-server-amd64.iso

Server Software: Apache, Mysql, PHP

RAM:1 GB

CPU: minimum 1 Ghz CPU

Harddrive: 15 GB for OS and min 1 GB for DMS System ++ estim. 0.3MB per Page

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

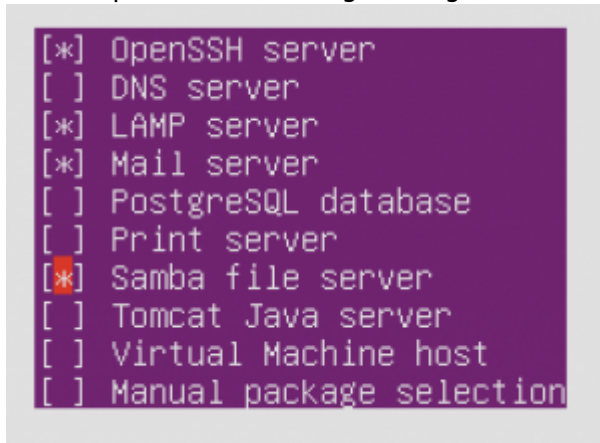
<http://beamweb.de/postsorter/doku.php?id=dms:requirements>

Last update: **2015/11/13 22:08**



# Packages to Install

Install Ubuntu 14.04 LTS server edition with the installer in default mode.  
On Setup choose following Packages:



Setup Question Mysql:  
Choose your Mysql Password

Setup Mail /Postfix  
Choose: Internet-Site  
Name: example:dms-test.lcl  
SSL-Name dms-test.lcl

Allow Root to Access Remote via SSH [Root Zugriff erlauben Ubuntu > 14.04](#)

Disable IPv6 if apt-get update wants to connect via IPv6 and Gateway is not Available [IPv6 deaktivieren in Linux](#)

After Installation the System requires about 1.4 GB of Disk Space

```
root@DMS-TEST:~# df -h
Dateisystem    Größe Benutzt Verf. Verw% Eingehängt auf
/dev/sda1      19G   1,4G   16G    9% /
none           4,0K     0   4,0K    0% /sys/fs/cgroup
udev           991M   4,0K  991M    1% /dev
tmpfs          201M   1,1M  200M    1% /run
none           5,0M     0   5,0M    0% /run/lock
none          1001M     0 1001M    0% /run/shm
none           100M     0   100M    0% /run/user
```

OPTIONAL: Install Webmin [Recommendation](#)

## Install required Packages

Needed by OCRmyPDF:

```
ImageMagick
GNU Parallel
poppler-utils
unpaper
tesseract
tesseract-data
lxml //python
reportlab //python
ghostscript
java
```

```
apt-get install imagemagick parallel poppler-utils unpaper tesseract-ocr
tesseract-ocr-deu tesseract-ocr-eng python-lxml python-reportlab
ghostscript openjdk-7-jre-headless
```

After sending the Command some dependencies are solved and installed:

```
root@DMS-TEST:~# apt-get install imagemagick parallel poppler-utils unpaper
tesseract-ocr tesseract-ocr-deu tesseract-ocr-eng python-lxml python-
reportlab ghostscript openjdk-7-jre-headless
Paketlisten werden gelesen... Fertig
Abhängigkeitsbaum wird aufgebaut.
Statusinformationen werden eingelesen.... Fertig
Die folgenden zusätzlichen Pakete werden installiert:
  ca-certificates-java fontconfig fontconfig-config fonts-dejavu-core
gsfonts
  hicolor-icon-theme imagemagick-common java-common libart-2.0-2 libasyncns0
  libcairo2 libcroco3 libcupsfilters1 libcupsimage2 libdatriel
  libdjvulibre-text libdjvulibre21 libfftw3-double3 libflac8 libfontconfig1
  libgdk-pixbuf2.0-0 libgdk-pixbuf2.0-common libgif4 libgomp1 libgraphite2-3
  libgs9 libgs9-common libharfbuzz0b libijs-0.35 libilmbase6 libjasper1
  libjbig0 libjbig2dec0 libjpeg-turbo8 libjpeg8 liblcms2-2 liblpt4
liblqr-1-0
  libltdl7 libmagickcore5 libmagickcore5-extra libmagickwand5 libnetpbm10
  libnspr4 libnss3 libnss3-nssdb libogg0 libopenexr6 libpango-1.0-0
  libpangocairo-1.0-0 libpangoft2-1.0-0 libpaper-utils libpaper1
libpixman-1-0
  libpoppler44 libpulse0 librsvg2-2 librsvg2-common libsctp1 libsndfile1
  libtesseract3 libthai-data libthai0 libtiff5 libvorbis0a libvorbisenc2
  libwebp5 libwebpmux1 libwmf0.2-7 libxcb-render0 libxcb-shm0 libxrender1
  libxslt1.1 lksctp-tools netpbm poppler-data python-pil python-renderpm
  python-reportlab-accel tesseract-ocr-equ tesseract-ocr-osd tzdata
  tzdata-java
Vorgeschlagene Pakete:
  ghostscript-x hpijs imagemagick-doc autotrace cups-bsd lpr lprng enscript
```

```

ffmpeg gimp gnuplot grads hp2xx html2ps libwmf-bin mplayer povray radiance
sane-utils texlive-base-bin transfig xdg-utils ufraw-batch default-jre
equivs libfftw3-bin libfftw3-dev fonts-droid libjasper-runtime
liblcms2-utils ttf-baekmuk ttf-arphic-gbsn00lp ttf-arphic-bsmi00lp
ttf-arphic-gkai00mp ttf-arphic-bkai00mp pulseaudio librsvg2-bin
libwmf0.2-7-gtk icedtea-7-jre-jamvm libnss-mdns sun-java6-fonts
fonts-dejavu-extra fonts-ipafont-gothic fonts-ipafont-mincho
ttf-wqy-microhei ttf-wqy-zenhei ttf-indic-fonts-core ttf-telugu-fonts
ttf-oriya-fonts ttf-kannada-fonts ttf-bengali-fonts fonts-arphic-ukai
fonts-arphic-uming fonts-unfonts-core python-lxml-dbg python-pil-doc
python-pil-dbg python-renderpm-dbg pdf-viewer python-egenix-mxtexttools
python-reportlab-doc

```

Die folgenden NEUEN Pakete werden installiert:

```

ca-certificates-java fontconfig fontconfig-config fonts-dejavu-core
ghostscript gsfonts hicolor-icon-theme imagemagick imagemagick-common
java-common libart-2.0-2 libasyncns0 libcairo2 libcroco3 libcupsfilters1
libcupsimage2 libdatriel libdjvulibre-text libdjvulibre21 libfftw3-double3
libflac8 libfontconfig1 libgdk-pixbuf2.0-0 libgdk-pixbuf2.0-common libgif4
libgomp1 libgraphite2-3 libgs9 libgs9-common libharfbuzz0b libijs-0.35
libilmbase6 libjasper1 libjbig0 libjbig2dec0 libjpeg-turbo8 libjpeg8
liblcms2-2 libleft4 liblqr-1-0 libltdl7 libmagickcore5 libmagickcore5-

```

extra

```

libmagickwand5 libnetpbm10 libnspr4 libnss3 libnss3-nssdb libogg0
libopenexr6 libpango-1.0-0 libpangocairo-1.0-0 libpangoft2-1.0-0
libpaper-utils libpaper1 libpixmap-1-0 libpoppler44 libpulse0 librsvg2-2
librsvg2-common libsctp1 libsndfile1 libtesseract3 libthai-data libthai0
libtiff5 libvorbis0a libvorbisenc2 libwebp5 libwebpmux1 libwmf0.2-7
libxcb-render0 libxcb-shm0 libxrender1 libxslt1.1 lksctp-tools netpbm
openjdk-7-jre-headless parallel poppler-data poppler-utils python-lxml
python-pil python-renderpm python-reportlab python-reportlab-accel
tesseract-ocr tesseract-ocr-deu tesseract-ocr-eng tesseract-ocr-equ
tesseract-ocr-osd tzdata-java unpaper

```

Die folgenden Pakete werden aktualisiert (Upgrade):

tzdata

1 aktualisiert, 93 neu installiert, 0 zu entfernen und 94 nicht aktualisiert.

Es müssen 92,8 MB an Archiven heruntergeladen werden.

Nach dieser Operation werden 229 MB Plattenplatz zusätzlich benutzt.

Needed by OCR.Script:

```

lockfile
convmv
zbarimg
pdftk

```

**apt-get install lockfile-progs convmv zbar-tools pdftk**

There are also some dependencies thats been solved

```
root@DMS-TEST:~# apt-get install lockfile-progs convmv libbarcode-zbar-perl
```

```
pdftk
Paketlisten werden gelesen... Fertig
Abhängigkeitsbaum wird aufgebaut.
Statusinformationen werden eingelesen.... Fertig
lockfile-progs ist schon die neueste Version.
Die folgenden zusätzlichen Pakete werden installiert:
  gcj-4.8-jre-lib libasound2 libasound2-data libgcj-common libgcj14 libv4l-0
  libv4lconvert0 libzbar0
Vorgeschlagene Pakete:
  libasound2-plugins alsa-utils libgcj14-dbg libgcj14-awt
Die folgenden NEUEN Pakete werden installiert:
  convmv gcj-4.8-jre-lib libasound2 libasound2-data libbarcode-zbar-perl
  libgcj-common libgcj14 libv4l-0 libv4lconvert0 libzbar0 pdftk
0 aktualisiert, 11 neu installiert, 0 zu entfernen und 94 nicht
aktualisiert.
Es müssen 21,4 MB an Archiven heruntergeladen werden.
Nach dieser Operation werden 72,8 MB Plattenplatz zusätzlich benutzt.
```

## Add User for DMS

Lets add a User for the DMS

```
root@DMS-TEST:/etc# adduser dms-test
Lege Benutzer »dms-test« an ...
Lege neue Gruppe »dms-test« (1001) an ...
Lege neuen Benutzer »dms-test« (1001) mit Gruppe »dms-test« an ...
Erstelle Home-Verzeichnis »/home/dms-test« ...
Kopiere Dateien aus »/etc/skel« ...
Geben Sie ein neues UNIX-Passwort ein:
Geben Sie das neue UNIX-Passwort erneut ein:
passwd: password updated successfully
Changing the user information for dms-test
Enter the new value, or press ENTER for the default
  Full Name []: DMS-Test
  Room Number []:
  Work Phone []:
  Home Phone []:
  Other []:
Sind diese Informationen korrekt? [J/n] j
root@DMS-TEST:/etc#
```

.. and add it to the www-data group

```
root@DMS-TEST:/etc# adduser dms-test www-data
Füge Benutzer »dms-test« der Gruppe »www-data« hinzu ...
Adding user dms-test to group www-data
Fertig.
root@DMS-TEST:/etc#
```

## Now create a MySQL Database

```
mysql -u root -p<password>
```

```
mysql> create database dms;  
Query OK, 1 row affected (0.00 sec)
```

Add a User for DB Access

```
mysql> CREATE USER 'dms'@'localhost' IDENTIFIED BY 'dmstest';  
Query OK, 0 rows affected (0.00 sec)
```

Grant Privileges to the User for Database dms

```
mysql> use dms  
Database changed  
mysql> GRANT ALL PRIVILEGES ON dms.* TO 'dms'@'localhost';  
Query OK, 0 rows affected (0.00 sec)
```

Check if you have access:

```
mysql> show grants for dms@localhost;  
+-----+  
-----+  
| Grants for dms@localhost  
|  
+-----+  
-----+  
| GRANT USAGE ON *.* TO 'dms'@'localhost' IDENTIFIED BY PASSWORD  
'*D26F4EC699CF976CF5F6EB655C796E70B6F76340' |  
| GRANT ALL PRIVILEGES ON `dms`.* TO 'dms'@'localhost'  
|  
+-----+  
-----+  
2 rows in set (0.00 sec)
```

## Import the Database Structure

```
dms@DMS-TEST:~$ mysql -u dms -pdmstest dms <postsorter.sql  
dms@DMS-TEST:~$
```

## Copy Files and Scripts to the Webserver

You can use WIN-SCP if you use Windows.  
Use the User dms-test for this.

Before you can copy the Files to the Webservers Directory you have to allow the www-data group the Permission to do so.

you have already add the User **dms-test** to the www-data Group.

```
root@DMS-TEST: cd /var/www/  
root@DMS-TEST:/var/www# chgrp www-data html/  
root@DMS-TEST:/var/www# chmod g+w html/
```

Check if you can reach the Webserver

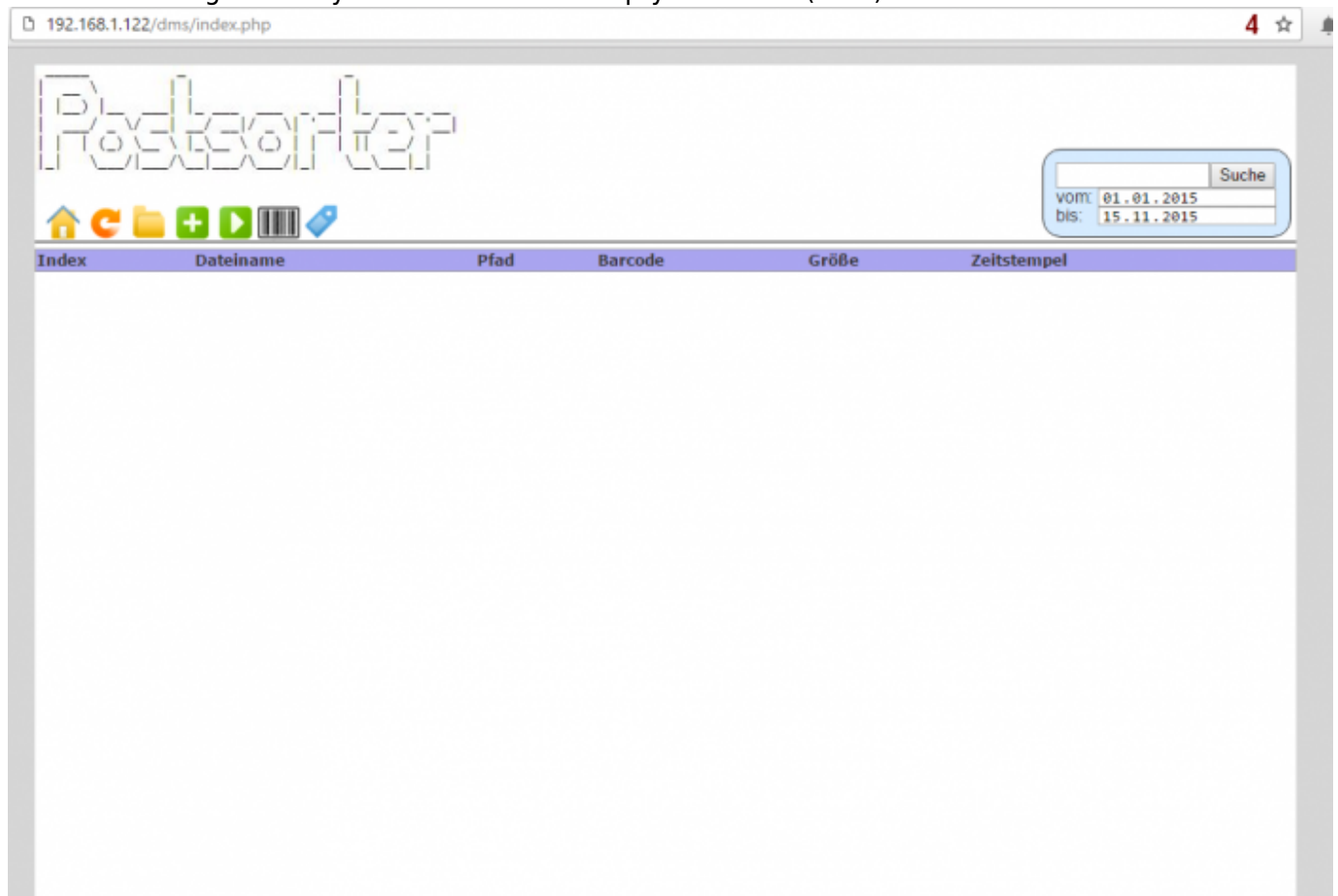
<http://IP-OF-YOUR-SERVER/dms>

Maybe you see :

**Verbindungsversuch zur Datenbank postsorter fehlgeschlagen**

So you have to correct the Database Settings in */dms/config.php*

If the DB Settings are ok you should see the Empty Postsorter(DMS) Web-Interface



### Set the Permissions for the Scripts and Folders

```
cd /var/www/html/dms/  
chmod u+x run_ocr.sh OCRmyPDF.sh  
chmod u+x src/ocrPage.sh
```

If you got an Error on Adding Files to the Mysql Table pdftext  
Please disable your apparmor profile for the Mysql Service,  
[disable\\_apparmor](#).

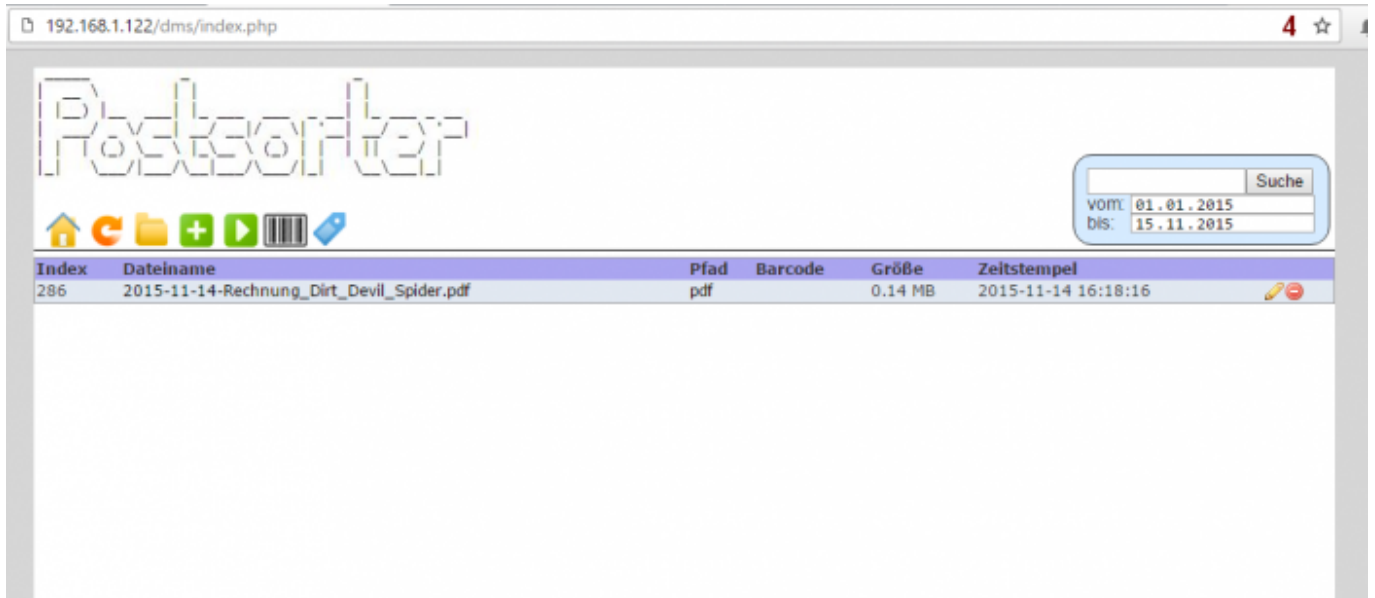
### Let the Magic begin

To Test the Postsorter copy a PDF or JPG file to the /var/www/html/dms/scan folder

run the Script from Commandline:

```
dms-test@DMS-TEST:cd /var/www/html/dms/  
dms-test@DMS-TEST:./run_ocr.sh  
Can't rename scan/*.PDF scan/*.pdf: Datei oder Verzeichnis nicht gefunden  
Ready!  
OCRmyPDF fuer Rechnung_Dirt_Devil_Spider.pdf startet..  
Kein Barcode gefunden  
  
Kein Schlagworttreffer, Pdf nur mit Datum abgespeichert...  
Kein neues jpg  
dms-test@DMS-TEST:/var/www/html/dms$
```

Result:



From:  
<http://beamweb.de/postsorter/> - **Doku**

Permanent link:  
<http://beamweb.de/postsorter/doku.php?id=dms:preinstall>

Last update: **2015/11/14 16:21**







# Recommendation

To simplify the Administration of the System i prefer the use of Webmin.

```
# apt-get update
# apt-get install libnet-ssleay-perl libauthen-pam-perl libio-pty-perl apt-
show-versions
```

```
# wget http://prdownloads.sourceforge.net/webadmin/webmin_1.7x_all.deb
```

```
# dpkg --install webmin_1.7x_all.deb
```

```
root@DMS-TEST:~# dpkg --install webmin_1.770_all.deb
Vormals nicht ausgewähltes Paket webmin wird gewählt.
(Lese Datenbank ... 60369 Dateien und Verzeichnisse sind derzeit
installiert.)
Vorbereitung zum Entpacken von webmin_1.770_all.deb ...
Entpacken von webmin (1.770) ... ////////// The extract takes some Time,
webmin (1.770) wird eingerichtet ...
Webmin install complete. You can now login to https://DMS-TEST:10000/
as root with your root password, or as any user who can use sudo
to run commands as root.
Trigger für ureadahead (0.100.0-16) werden verarbeitet ...
ureadahead will be reprofiled on next reboot
```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

<http://beamweb.de/postsorter/doku.php?id=dms:recommendation>

Last update: **2015/11/13 23:54**



# Configuration

## Database

Please change your Database configuration in the config Files

```
/var/www/html/dms/config/config.ini.php
```

and

```
/var/www/html/dms/run_ocr.sh
```

You can set the API Key for NotifyMyAndroid in the

```
/var/www/html/dms/run_ocr.sh
```

file. Line 120 and 134

## Cron Job

add a cronjob to run the script crontab -e

```
10,20,30,40,50 * * * * cd /var/www/html/dms && ./run_ocr.sh >/dev/null 2>&1  
#Run OCR
```

## Wrapper Config

To start indexing from the Web Frontend there is a wrapper to run the php file. you have to set the Permissions for this file to

```
chmod u+s php_run_ocr
```

```
rwsr-xr-x 1 dms-test www-data 7382 Apr 7 2015 php_run_ocr
```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

<http://beamweb.de/postsorter/doku.php?id=dms:configuration>

Last update: **2016/01/05 00:16**



# Help Content

## Fulltext Search Operators

A phrase that is enclosed within double quote (“,”) characters matches only rows that contain the phrase literally, as it was typed. The full-text engine splits the phrase into words and performs a search in the FULLTEXT index for the words. Nonword characters need not be matched exactly: Phrase searching requires only that matches contain exactly the same words as the phrase and in the same order. For example, „test phrase“ matches „test, phrase“.

If the phrase contains no words that are in the index, the result is empty. The words might not be in the index because of a combination of factors: if they do not exist in the text, are stopwords, or are shorter than the minimum length of indexed words.

The following examples demonstrate some search strings that use boolean full-text operators:

```
'apple banana'
```

Find rows that contain at least one of the two words.

```
'+apple +juice'
```

Find rows that contain both words.

```
'+apple macintosh'
```

Find rows that contain the word “apple”, but rank rows higher if they also contain “macintosh”.

```
'+apple -macintosh'
```

Find rows that contain the word “apple” but not “macintosh”.

```
'+apple ~macintosh'
```

Find rows that contain the word “apple”, but if the row also contains the word “macintosh”, rate it lower than if row does not. This is “softer” than a search for '+apple -macintosh', for which the presence of “macintosh” causes the row not to be returned at all.

```
'+apple +(>turnover <strudel)'
```

Find rows that contain the words “apple” and “turnover”, or “apple” and “strudel” (in any order), but rank “apple turnover” higher than “apple strudel”.

```
'apple*'
```

Find rows that contain words such as “apple”, “apples”, “applesauce”, or “applet”.

```
' "some words" '
```

## Barcode Tags

You can scan a Separator Page with a Barcode, if the Barcode is found during indexing the Scanned File is named to the Barcode content.

Also a Folder is created (Barcode Content is the Folder Name) and the file is copied in it. The First Page (The Separator Page) would be removed after indexing and doesn't appear in the Document.

If you have Scans with Barcode on the Page (e.g. some Insurances, Phone Companies ..) and there is no separator Page, please be sure you make the Barcode unreadable with a Marker/Pen. If not the first Page (with the Barcode) is removed.

Barcodes can be created within the Web Frontend.

## Tag Words

You can add tag Words for sorting your documents, if a tag word is found during indexing the File name would be changed to the given Tag.

This happens if NO Barcode is found.

Tags can be created/changed through the Frontend.

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

<http://beamweb.de/postsorter/doku.php?id=dms:help>

Last update: **2016/01/04 23:58**



# Mobile Scanning/Upload

You can scan and Upload Images to the Postsorter to index Files like Cassier Bons Receipts and so on.

- Scan the Files with your Phone Camera
- Install the App [AndFTP](#)
- configure the Access to the DMS Server IP , Port and User
- Share your Photo with And FTP

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

[http://beamweb.de/postsorter/doku.php?id=dms:mobile\\_upload](http://beamweb.de/postsorter/doku.php?id=dms:mobile_upload)

Last update: **2016/01/04 22:50**



# DMS Backup

## DB-Backup

[db\\_backup.sh](#)

```
#!/bin/bash

date=`date +%Y-%m-%d`
USER="root"
PASSWORD="xyz"
db="postsorter"
OUTPUT="/var/www/html/dms/DB-Backup"
mysqldump --force --opt --user=$USER --password=$PASSWORD --databases
$db > $OUTPUT/$date.$db.sql
      gzip $OUTPUT/$date.$db.sql
```

## Ordner Backup per rsync

[backup.sh](#)

```
#!/bin/sh

date=`date +%Y-%m-%d`
USER="postsorter"
PASSWORD="pass"
db="postsorter"
OUTPUT="/var/www/html/dms/DB-Backup"

mysqldump --force --opt --user=$USER --password=$PASSWORD --databases
$db > $OUTPUT/$date.$db.sql
      gzip $OUTPUT/$date.$db.sql

rsync -ave ssh /var/www/html/dms
admin@192.168.1.170:/share/Private/Dokumente/Scan_Backup/

errorcode=$?

if [ $errorcode != "0" ]
then
    echo "Backup QNAPfehgeschlagen mit Fehler:$errorcode"
    wget -q -O /dev/null
    'http://www.notifymyandroid.com/publicapi/notify?apikey=yourkey&applica
```

```
tion=Backup&event=Backup&description=Backup Scanfiles nach TS-251
fehlgeschlagen mit Fehler: '$errorcode'.&priority=2'
    exit 1
else
    echo "Backup erfolgreich QNAP"
    wget -q -O /dev/null
    'http://www.notifymyandroid.com/publicapi/notify?apikey=yourkey&applica
tion=Backup&event=Backup&description=Backup Scanfiles nach TS-251
erfolgreich.&priority=2'

fi

## owncloudcmd /var/www/html/dms/
owncloud://user:pass@cloud.xyz.de/remote.php/webdav/Scan_Backup
```

From:  
<http://beamweb.de/postsorter/> - **Doku**

Permanent link:  
[http://beamweb.de/postsorter/doku.php?id=dms:scripts:backup\\_script](http://beamweb.de/postsorter/doku.php?id=dms:scripts:backup_script)

Last update: **2016/01/04 22:37**





# Postsorter Run Script

The **run\_ocr.sh** is the Main Script, this is executed via Button in the Web Frontend (and the Wrapper Script **php\_run\_ocr**) or via cron Job, i recommend a period of 10 Minutes for slower Computers because if you scan many pages it can take some time to finish the job.

[run\\_ocr.sh](#)

```
#!/bin/bash
#####
#                               #
# Postsortierer by Oli          #
#                               #
#####

#### Konfig Parameter

date=`date +%Y-%m-%d`          # Filedatums Format
pwd=`pwd`                      # aktueller Script Pfad
pdfdir=pdf                    # PDF Stammverzeichnis
scandir=scan                  # Upload Verzeichnis /Hotfolder freizugeben per
                             # bsw. FTP
thumbsize="x400"             # Vorschaubildgrösse
dbuser=postsorter            # Datenbankuser
dbpass=xyz                   # DB Passwort
db=postsorter                # Datenbank

#Nur ein Script auf einmal, gegen konkurrierende zugriffe schützen falls
#der Cronjob mal länger dauert

if [ -e "ocrlock.lock" ]      #Prüfen ob
Filelock besteht
then
    exit 1
fi

    lockfile-create ocrlock    #Lockfile erstellen
countscandir=`ls -l $scandir/*.pdf 2>/dev/null | wc -l`          #PDFs
in scandir Zählen
countscriptdir=`ls -l *.pdf 2>/dev/null | wc -l`                #PDFs in
Zählen
countscandirb=`ls -l $scandir/*.PDF 2>/dev/null | wc -l`       #PDFs in
scandir Zählen
countscriptdirb=`ls -l *.PDF 2>/dev/null | wc -l`               #PDFs in
Zählen
```

```

if [ $countscandir != ] || [ $countscriptdir != ] || [ $countscandirb
!= ] || [ $countscriptdirb != ]      # Nur ausführen wenn überhaupt
ein PDF im scanordner ist.
then                                  #..

    rename 's/\.PDF$/\.pdf/' $scandir/*.PDF      #
    großgeschriebene Dateiendungen nach klein Umschreiben
    convmv --notest -f iso-8859-1 -t UTF-8 -r "$scandir"
# File Encoding erzwingen nach UTF-8 , zwecks umlauten
    mv $scandir/*.* .                          #Dateien aus dem
Scandir ins scriptdir verschieben

for i in *.pdf; do                      # Schleife für jede PDF
im Verzeichnis starten
lockfile-touch -o ocrlock                # Filelock Datei
erneuern falls das Script länger läuft

echo "OCRmyPDF fuer $i startet.."

    $pwd/OCRmyPDF.sh -f -l deu "$i" "ocr-$i"
#Externes OCR Script starten

    groesse=`stat --format="%s" "ocr-$i"`
#Dateigröße herrausfinden und in String speichern
    convert -density 150 "$i[0]" -quality 100 -sharpen 0x1.0 "$i.jpg"
# JPG erzeugen um möglichen Barcode zu suchen
    barcode=`zbarimg -q --raw "$i.jpg"`        # Barcode
suchen und in Variable speichern
    rm "$i.jpg"                                # Bild wieder
löschen

if [ -z "$barcode" ];then                #falls Barcode
nicht gefunden...

echo "Kein Barcode gefunden"

    x="ocr-$i"                                # Dsteiendung
abschneiden
    y="{x%.pdf}"
    z="{y##*/}"

    pdftotext "ocr-$i"                        #Text aus PDF holen
und Schlagwort zu suchen und um die DB zu befüllen

    string=`cat "$z.txt"`                      #txt Datei in
String lesen

###Stringsuche nach Schlagwörtern

```

```
while read benutzer benennung schlagworte; do

while IFS=':' read -ra SW; do

countall=$(grep -o ":" <<< "$schlagworte" | wc -l)
countall=$((countall + 1))
#echo "Anzahl Schlagworte: "$countall

    for b in "${SW[@]}"; do
#echo $b

        if [[ $string == *"$b"* ]]
            then
                count=$((count + 1))

            fi
        done

#echo "Anzahl Treffer: "$count
        if [[ $count == $countall ]]
            then
#echo "Alle Schlagworte gefunden"

name=$benennung
folder=$benennung
        fi
        done <<< "$schlagworte"

unset count
unset countall

#echo "$benutzer,$benennung,$schlagworte"

done < <(/usr/bin/mysql -u $dbuser -p$dbpass -D $db --default-
character-set=utf8 --skip-column-names --ignore-spaces -e "SELECT
benutzer, benennung, schlagworte FROM schlagworte ")

echo $name
echo $folder

        convert -thumbnail $thumbsize "ocr-$i[0]" "$i.png" #
PDF Vorschaubild erstellen

        if [ -z "$folder" ];then #wenn kein
Schlagwort übereinstimmt PDF nur mit Datum Speichern
            mv "ocr-i" "$pdfdir/$date-i" # PDF
verschieben
```

```

    mv "$i.png" "$pdfdir/$date-$i.png" #
Vorschaubild verschieben

### Daten in die MYSQL Datenbank speichern

    echo "INSERT INTO files
(name,pfad,vorschau,groesse,pdftext,barcode,original,datetime) VALUES
('$date-$i','$pdfdir','$pdfdir/$date-
$i.png','$groesse',LOAD_FILE('$pwd/$z.txt'),'barcode','$i',NOW() );" |
/usr/bin/mysql -u $dbuser -p$dbpass -D $db --default-character-
set=utf8;
wget -q -O /dev/null
"http://www.notifymyandroid.com/publicapi/notify?apikey=xyz&application
=PostScan&event=Scan&description=PDF $i nach $pdfdir
gescannt.&priority=1"
echo "Kein Schlagworttreffer, Pdf nur mit Datum abgespeichert... "

else # Wenn Schlagwort
gefunden dann Nae und Ordner ändern

    mkdir -p "$pdfdir/$folder" #
Verzeichnis erstellen wenns noch nicht da ist
    chgrp www-data "$pdfdir/$folder"
    chmod g+w "$pdfdir/$folder"
    mv "ocr-$i" "$pdfdir/$folder/$date-$name-$i" #
Dateiverschieben
    mv "$i.png" "$pdfdir/$folder/$date-$name-$i.png" #
Vorschaubild verschieben

    echo "INSERT INTO files
(name,pfad,vorschau,groesse,pdftext,barcode,original,datetime) VALUES
('$date-$name-$i','$pdfdir/$folder','$pdfdir/$folder/$date-$name-
$i.png','$groesse',LOAD_FILE('$pwd/$z.txt'),'barcode','$i',NOW() );" |
/usr/bin/mysql -u $dbuser -p$dbpass -D $db --default-character-
set=utf8;

    wget -q -O /dev/null
"http://www.notifymyandroid.com/publicapi/notify?apikey=xyz&application
=PostScan&event=Scan&description=PDF $i nach $pdfdir/$folder
gescannt.&priority=1" # Notifikation ans Telefon

echo "Schlagworttreffer, PDF einsortiert.."

fi ## Benennungs FI für Kein Barcode ENDE

### Barcodeteil falls in der Trennseite ein Barcode ist

```

```

else                                                    ## Barcode gefunden Else

    pdftotext "ocr-$i"                                # Text aus PDF
    holen für DB Eintrag
    x="ocr-$i"
    y="{x%.pdf}"                                       #Dateiendung
    abschneiden für DB eintrag
    z="{y##*/}"

echo $barcode                                         #Barcode ausgeben
    mv "ocr-$i" "cut-$i"                             # Datei
umbenennen fuers entfernen der ersten Seite

    fin="$pwd/cut-$i"                                   # Pfad
Workaround für pdftk
    fout="$pwd/ocr-$i"                                  # Workaround für
pdftk

anzahlseiten=$(/usr/bin/pdftk "$fin" dump_data|grep
NumberOfPages| awk '{print $2}')                       #Anzahl PDF Seiten herausfinden

    if [[ $anzahlseiten > 1 ]]
    then
        /usr/bin/pdftk "$fin" cat 2-end output "$fout"
# Erste Seite wegschneiden falls barcode gefunden, Erste Seite gleich
Trennseite, wenns nur eine Seite gibt lassen nichts tun
    else
        cp "cut-$i" "ocr-$i"
#zurückkopieren falls nur eine Seite und kopie damit der aufräumjob
passt
    fi

    convert -thumbnail $thumbsize "ocr-$i[0]" "$i.png" #
Vorschaubild erzeugen

    mkdir -p "$pdfdir/$barcode"                       # Falls
ein neuer Barcode kommt Verzeichnis anlegen wenns noch nicht da ist
# chgrp -R www-data "$pdfdir"
# chmod -R 775 "$pdfdir"
    mv "ocr-$i" "$pdfdir/$barcode/$date-$barcode-$i"
# Fertige PDF wegsortieren
    mv "$i.png" "$pdfdir/$barcode/$date-$barcode-$i.png"
# Vorschaubild verschieben

```

```
chgrp -R www-data "$pdfdir"
chmod -R 775 "$pdfdir"
rm "cut-$i" # Aufräumen

#####Daten in SQL schreiben

echo "INSERT INTO files
(name,pfad,vorschau,groesse,pdftext,barcode,original,datetime) VALUES
('$date-$barcode-$i','$pdfdir/$barcode','$pdfdir/$barcode/$date-
$i.png','$groesse',LOAD_FILE('$pwd/$z.txt'),'$barcode','$i',NOW() );" |
/usr/bin/mysql -u $dbuser -p$dbpass -D$db --default-character-set=utf8;

## wget -q -O /dev/null
"http://www.notifymyandroid.com/publicapi/notify?apikey=5bc537694fcc4fe
14805ff39e3bd1536f4397d821a2c3ad3&application=PostScan&event=Scan&descr
iption=PDF $i nach $pdfdir/$barcode gescannt.&priority=1" #
Notifikation ans Telefon

SUBJECT="Scan-Job"
EMAIL="oliver.amann@gmail.com"
EMAILMESSAGE="/tmp/emailmessage.txt"
echo "Hallo _$Benutzer_"> $EMAILMESSAGE
echo "PDF $i nach $pdfdir/$barcode gescannt." >>$EMAILMESSAGE
/usr/bin/mail -s "$SUBJECT" "$EMAIL" < $EMAILMESSAGE

echo "Barcode gefunden und zur Benennung genutzt PDF verschoben DB
geschrieben "

fi #Barcode gefunden ja/nein FI

###Abschlussarbeiten

rm "$z.txt" #OCR Textdatei löschen
mkdir -p "$pdfdir/originale"
#Backupverzeichnis für Originale erstellen
chgrp www-data "$pdfdir/originale"
chmod g+w "$pdfdir/originale"
mv "$i" "$pdfdir/originale/" #Originale
Verschieben
chgrp -R www-data $pdfdir/originale
chmod -R 775 $pdfdir/originale
unset name
unset folder
unset barcode
```

```

unset groesse

done                                     #Schleifen Ende nächstes PDF
else                                     # Else für Check ob PDF da
ist
echo "Kein neues PDF"

fi                                       #End if für Check ob PDF da ist

countscandirjpg=`ls -l $scandir/*.jpg 2>/dev/null | wc -l`
#JPG Zählen
countscriptdirjpg=`ls -l *.jpg 2>/dev/null | wc -l`           #JPG
Zählen
countscandirjpgb=`ls -l $scandir/*.JPG 2>/dev/null | wc -l`
#JPG Zählen
countscriptdirjpgb=`ls -l *.JPG 2>/dev/null | wc -l`         #JPG
Zählen

    if [ $countscandirjpg != ] || [ $countscriptdirjpg != ] || [
$countscandirjpgb != ] || [ $countscriptdirjpgb != ] # Nur
ausführen wenn überhaupt ein jpg im scanordner ist.
    then
# Nur ausführen wenn überhaupt ein jpg im scanordner ist.
        rename 's/\.JPG$/\.jpg/' $scandir/*.JPG             #
großgeschriebene Dateiendungen nach klein Umschreiben
        convmv --notest -f iso-8859-1 -t UTF-8 -r "$scandir" #
File Encoding erzwingen nach UTF-8 , zwecks umlauten

        mv $scandir/*. * .                                # Alles aus dem
Scandir ins Scriptdir schieben

        for j in *.jpg; do

            convert "$j" "$j.pdf"                          # Jpg nach PDF
konvertieren
            rm "$j"                                         # JPG löschen
done

        else                                               # Else für Check ob JPG da
ist
echo "Kein neues jpg"
fi                                                         # JPG Bearbeitung Ende

lockfile-remove ocrlock                                # Alles gut? dann
Filelock entfernen

```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

[http://beamweb.de/postsorter/doku.php?id=dms:scripts:run\\_ocr](http://beamweb.de/postsorter/doku.php?id=dms:scripts:run_ocr)

Last update: **2016/01/05 00:06**





# PHP Run Wrapper

Damit man das Script per PHP/Webseite ausführen kann benötigt man ein „Programm“

[wrapper.c](#)

```
#include <stdlib.h>
#include <sys/types.h>
#include <unistd.h>

int
main (int argc, char *argv[])
{
    setuid ();

    /* WARNING: Only use an absolute path to the script to execute,
     *           a malicious user might fool the binary and execute
     *           arbitrary commands if not.
     *           zum kompilieren einfach: gcc wrapper.c -o php_run_ocr
     verwenden
     * */
    char cwd[1024];
    system ("/bin/sh /var/www/html/dms/php_run_ocr.sh");

    return ;
}
```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

<http://beamweb.de/postsorter/doku.php?id=dms:scripts:wrapper.c>

Last update: **2015/11/13 22:10**



# Nur MySQL Structur ausgeben

Syntax:

```
mysqldump -d -h localhost -u root -pmypassword databasename > dumpfile.sql
```

The only option that is different than creating an entire backup is the -d switch, which tells mysqldump not to output the data.

Example:

```
mysqldump -d -h localhost -u root -p2Uad7as9 database01 > dumpfile.sql
```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

[http://beamweb.de/postsorter/doku.php?id=konfig:backup\\_mysql\\_structure\\_only](http://beamweb.de/postsorter/doku.php?id=konfig:backup_mysql_structure_only)

Last update: **2015/11/13 23:52**



# AppArmor disable

Before:

```
root@DMS-TEST:~# apparmor_status
apparmor module is loaded.
5 profiles are loaded.
5 profiles are in enforce mode.
  /sbin/dhclient
  /usr/lib/NetworkManager/nm-dhcp-client.action
  /usr/lib/connman/scripts/dhclient-script
  /usr/sbin/mysqld
  /usr/sbin/tcpdump
profiles are in complain mode.
2 processes have profiles defined.
2 processes are in enforce mode.
  /sbin/dhclient (731)
  /usr/sbin/mysqld (1082)
processes are in complain mode.
processes are unconfined but have a profile defined.
```

```
root@DMS-TEST:/etc/apparmor.d# ln -s /etc/apparmor.d/usr.sbin.mysqld
/etc/apparmor.d/disable/
root@DMS-TEST:/etc/apparmor.d# apparmor_parser -R
/etc/apparmor.d/usr.sbin.mysqld
```

After:

```
root@DMS-TEST:/etc/apparmor.d# apparmor_status
apparmor module is loaded.
4 profiles are loaded.
4 profiles are in enforce mode.
  /sbin/dhclient
  /usr/lib/NetworkManager/nm-dhcp-client.action
  /usr/lib/connman/scripts/dhclient-script
  /usr/sbin/tcpdump
profiles are in complain mode.
1 processes have profiles defined.
1 processes are in enforce mode.
  /sbin/dhclient (731)
processes are in complain mode.
processes are unconfined but have a profile defined.
```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

[http://beamweb.de/postsorter/doku.php?id=konfig:disable\\_apparmor](http://beamweb.de/postsorter/doku.php?id=konfig:disable_apparmor)

Last update: **2015/11/14 15:42**



# IPv6 deaktivieren in Linux

To disable ipv6, you have to open `/etc/sysctl.conf` using any text editor and insert the following lines at the end:

```
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv6.conf.default.disable_ipv6 = 1
net.ipv6.conf.lo.disable_ipv6 = 1
```

If ipv6 is still not disabled, then the problem is that `sysctl.conf` is still not activated.

To solve this, open a terminal(Ctrl+Alt+T) and type the command,

```
sudo sysctl -p
```

You will see this in the terminal:

```
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv6.conf.default.disable_ipv6 = 1
net.ipv6.conf.lo.disable_ipv6 = 1
```

After that, if you run:

```
$ cat /proc/sys/net/ipv6/conf/all/disable_ipv6
```

It will report:

```
1
```

If you see 1, ipv6 has been successfully disabled.

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

[http://beamweb.de/postsorter/doku.php?id=konfig:disable\\_ipv6](http://beamweb.de/postsorter/doku.php?id=konfig:disable_ipv6)

Last update: **2015/11/13 23:17**



# Root Zugriff erlauben Ubuntu > 14.04

Edit `/etc/ssh/sshd_config`, and **comment out** the following line:

```
PermitRootLogin without-password > #PermitRootLogin without-password
```

Just below it, add the following line:

```
PermitRootLogin yes
```

Then restart SSH:

```
service ssh restart
```

From:

<http://beamweb.de/postsorter/> - **Doku**

Permanent link:

[http://beamweb.de/postsorter/doku.php?id=konfig:ssh\\_allow\\_root\\_access](http://beamweb.de/postsorter/doku.php?id=konfig:ssh_allow_root_access)

Last update: **2015/11/13 23:30**

