Wikiprint Book

Title: Instalacja OCRa dla faktur w systemie eDokumenty

Subject: Archiwum - eDokumenty - elektroniczny system obiegu dokumentów, workflow i CRM - DeployerGuide/Customization/OCRInvoice

Version: 67

Date: 09/08/26 22:40:55

Table of Contents

Instalacja OCRa dla faktur w systemie eDokumenty *3*

Instalacja OCRa dla faktur w systemie eDokumenty

Poniższa instrukcja przedstawia uruchomienie mechanizmu OCRowania faktur w systemie eDokumenty działających na systemie Linux. Mechanizm jest obsługiwany od wersji 5.2.77.

Poniższa instrukcja została przygotowana na bazie systemu Linux Debian8

```
apt-get update
```

```
apt-get install autoconf-archive automake g++ libtool libleptonica-dev pkg-config
apt-get install git
apt-get install poppler-utils
apt-get install libjpeg-dev libtiff-dev libpng-dev
```

Jeśli pakiety leptonica 1.74+ nie są dostępne w dystrybucji w takim przypadku, konieczna będzie komplikacja ze źródeł

```
mkdir /usr/lib/leptonica
cd /usr/lib/leptonica
wget http://www.leptonica.org/source/leptonica-1.79.0.tar.gz
gunzip leptonica-1.79.0.tar.gz
tar -xf leptonica-1.79.0.tar
cd leptonica-1.79.0
./configure
make
make install
```

```
mkdir /usr/lib/tesseract
cd /usr/lib/tesseract
git clone https://github.com/tesseract-ocr/tesseract.git tesseract-ocr
cd tesseract-ocr/
./autogen.sh
./configure
make
make install
ldconfig
```

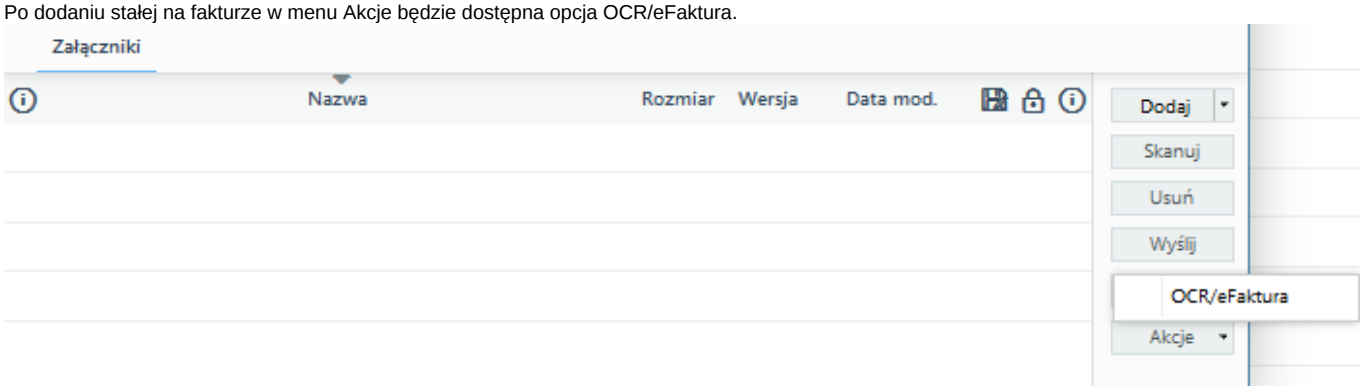
```
cd /usr/local/share/tessdata/
wget https://github.com/tesseract-ocr/tessdata_fast/raw/master/script/Latin.traineddata
wget https://github.com/tesseract-ocr/tessdata_fast/raw/master/pol.traineddata
wget https://github.com/tesseract-ocr/tessdata_fast/raw/master/eng.traineddata
wget https://github.com/tesseract-ocr/tessdata_fast/raw/master/osd.traineddata
```

Po pobraniu, zainstalowaniu oraz skompilowaniu pakietów ostatnim elementem jest dodanie stałej w config.inc domyślnie

```
vim /home/edokumenty/public_html/apps/edokumenty/config.inc
```

```
define('USE_NEW_OCR_FOR_EINVOICE', TRUE);
```

Po dodaniu stałej na fakturze w menu Akcje będzie dostępna opcja OCR/eFaktura.



Pakiety niezbędne do działania Bufora OCR

```
apt-get install rabbitmq-server
apt-get install supervisor
apt-get install python-opencv
apt-get install python-pip
pip install pika
apt install python-pil

# pdftk
apt-get install pdftk
```

Znane problemy:

1. Brak pakietu libpng12.so.0. W logach OCR pojawia się komunikat:

tesseract: error while loading shared libraries: libpng12.so.0: cannot open shared object file: No such file or directory

Sprawdzamy czy pakiet istnieje:

```
ls -ld $(locate -r libpng.*\.so.*)
```

Komenda powinna zwrócić nam:

```
lrwxrwxrwx 1 root root     19 kwi 18 22:12 /usr/lib/x86_64-linux-gnu/libpng16.so -> libpng16.so.16.28.0
lrwxrwxrwx 1 root root     19 kwi 18 22:12 /usr/lib/x86_64-linux-gnu/libpng16.so.16 -> libpng16.so.16.28.0
-rw-r--r-- 1 root root 206768 kwi 18 22:12 /usr/lib/x86_64-linux-gnu/libpng16.so.16.28.0
lrwxrwxrwx 1 root root     11 kwi 18 22:12 /usr/lib/x86_64-linux-gnu/libpng.so -> libpng16.so
```

Jeśli otrzymamy taką informację konieczne będzie ponowne kompilowanie leptonici oraz tesseract

Kompilowanie tesseract dla 1 wątku

```
./configure --disable-openmp
```