

Program do przetwarzania tekstu ze zdjęć

```
#=====określanie używanych modułów=====
=
from urllib.parse import urlparse
from pathlib import Path
import cv2
import pytesseract
import urllib.request
import os
from os import listdir
import pdfkit
path_wkhtmltopdf = Path("H:/wkhtmltopdf/bin/wkhtmltopdf.exe") #Wprowadzić trzeba ścieżkę do
    własnego
config = pdfkit.configuration(wkhtmltopdf=path_wkhtmltopdf)
from os.path import isfile, join
pytesseract.pytesseract.tesseract_cmd=r'H:\\teseract\\tesseract.exe' #Wprowadzić trzeba wła
    sną ścieżkę

#=====określanie ścieżek w zmiennych=====
=
_to_open=Path("text_from_image/links.txt")
_to_save=Path("text_from_image/images/")
_to_pdf=Path("text_from_image/pdf/")
_to_text=Path("text_from_image/txt/")
_txt_to_txt=Path("text_from_image/temp.txt")
_to_process=Path("text_from_image/To Process/")

#=====tworzenie inkrementujących folderów dla zdjęć=====
=
dirlist = [ item for item in os.listdir(_to_save) if os.path.isdir(os.path.join(_to_save,item
))]
dirlist.append("0")
_max_dir=max(int(item) for item in dirlist)
_max_dir=int(_max_dir)+1
_num=_max_dir
os.makedirs(f'{_to_save}\\{_num}',exist_ok=True)
_to_save=Path(f"text_from_image/images/{_num}/")
# =====znajdywanie najwyższego pdf numeru=====
#*****DOSTĘPNE, ALE CHWILOWO NIE UŻYWANE*****
onlyfiles = [f for f in listdir(_to_pdf) if isfile(join(_to_pdf,f))]
onlyfiles.append("0.pdf")
_max_file=max(int(item[:-4]) for item in onlyfiles)
_max_file=int(_max_file)+1
#=====znajdywanie najwyższego pliku txt numeru=====
onlyfiles_txt = [f for f in listdir(_to_text) if isfile(join(_to_text,f))]
onlyfiles_txt.append("0.txt")
_max_file_txt=max(int(item[:-4]) for item in onlyfiles_txt)
_max_file_txt=int(_max_file_txt)+1
#=====znajdywanie zdjęć do przetworzenia z folderu=====
to_process = [f for f in listdir(_to_process) if isfile(join(_to_process,f))]
to_process.insert(0,"*&%@#!.jpg")

#=====
links=[]
```

```
images=[]
```

```
def _get_links():
```

```
    try:
```

```
        f = open(_to_open)
```

```
        for line in f.readlines():
```

```
            if len(line)>3:
```

```
                link=urlparse(line)
```

```
                links.append(link.geturl())
```

```
    except FileNotFoundError:
```

```
        print("File [links.txt] not Found")
```

```
        f = open(_to_open, "w+")
```

```
    finally:
```

```
        if len(f.readlines())<1:
```

```
            _Clear_Session()
```

```
            print("File [links.txt] is empty")
```

```
            f.close()
```

```
            return False
```

```
        f.close()
```

```
        return True
```

```
def get_from_folder():
```

```
    if len(to_process)>1:
```

```
        for item in to_process[1:]:
```

```
            images.append(f'{_to_process}\\{item}')
```

```
    else:
```

```
        print("Folder is empty. Closing program")
```

```
        _Clear_Session()
```

```
def _Clear_Session():
```

```
    try:
```

```
        os.rmdir(f'{_to_save}')
```

```
        os.remove(f'{_to_text}\\{_max_file_txt}.txt')
```

```
    except:
```

```
        pass
```

```
def _download_image(url,fullname):
```

```
    urllib.request.urlretrieve(url,f'{_to_save}\\{fullname}\\')
```

```
def images_from_links():
```

```
    if _get_links():
```

```
        counter=0
```

```
        print("DOWNLOADING IMAGES")
```

```
        print("=====")
```

```
        for img in links:
```

```
            fullname=str(counter)+".jpg"
```

```
            _download_image(img,fullname)
```

```
            images.append(f'{_to_save}\\{fullname}')
```

```
            print(f'Progress: {round(((counter+1)/len(links))*100,2)}%')
```

```
            counter+=1
```

```
        print("=====")
```

```
#####Pozyskiwanie tekstu z zdjęć#####
```

```
def _txt_from_img(img):
```

```
    with open(f'{_to_text}\\{_max_file_txt}.txt','a+',encoding="utf-8") as f:
```

```
        f.write((pytesseract.image_to_string(img)))
```

```
def get_text():
```

```

try:
    print("PROCESSING IMAGES")
    print("=====")
    for i,img in enumerate(images):
        print(f"Progress: {round((i/len(images))*100,2)}%")
        _txt_from_img(img)
    print("=====")
except pytesseract.pytesseract.TesseractError:
    print("Image is not in valid format")
    _Clear_Session()

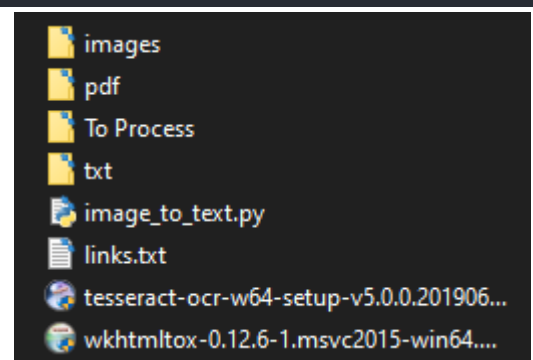
#=====NIE UŻYWANE. POTRZEBNE LEPSZE ROZWIĄZANIE NA ZACHOWANIE FORMATOWANIA
# TEKSTU PRZY ROBIENIU W PDF
def txt_to_txt():
    counter=0
    f = open(f'_{to_text}\\{_max_file_txt}.txt','r',encoding="utf-8")
    txt=open(_txt_to_txt,'a+',encoding="utf-8")
    for line in f.readlines():
        for i,j in enumerate(line):
            if counter>100:
                if j==" ":
                    txt.write("<br>")
                    counter=0
            txt.write(j)
            counter+=1
    f.close()
    txt.close()

"""
txt_to_txt()
"""

#CHWILAOWO UŻYTKOWE KOMENTY

# images_from_links()
# /
# get_from_folder()
#
# get_text()

```



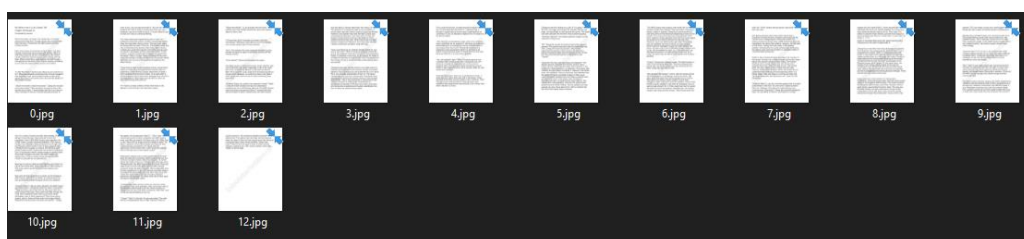
Program przyjmuje Zdjęcia i przetwarza z nich tekst następnie zapisując w pliku tekstowym. Zdjęcia można przetwarzać przez wrzucenie do folderu [To Process], Lub w pliku [links.txt]. Program obsługuje przetworzenie kilku zdjęć w plik Tekstowy. Przy podaniu linków zdjęcia są zapisywane w inkrementacyjnym folderze (FOLDERY: 1,2,3,4,5,...,n)[Program wyszukuje najwyższy folder i tworzy o jeden wyższy], a w nich pobrane zdjęcia od 0 do ilości zdjęć. Następnie zdjęcia są przetworzone i zapisane w pliku tekstowym o inkrementacyjnej nazwie.

Przykładowe działanie z linków:

- Plik links.txt

```
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-01.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-02.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-03.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-04.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-05.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-06.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-07.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-08.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-09.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-10.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-11.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-12.jpg?w=756
https://endlessfantasytranslations.files.wordpress.com/2016/12/chapter-199-131.jpg?w=756
```

- Pliki zostają pobrane z linku



- Zapisane w folderze (w tym przypadku [3])

```
DOWNLOADING IMAGES
=====
Progress: 7.69%
Progress: 15.38%
Progress: 23.08%
Progress: 30.77%
Progress: 38.46%
Progress: 46.15%
Progress: 53.85%
Progress: 61.54%
Progress: 69.23%
Progress: 76.92%
Progress: 84.62%
Progress: 92.31%
Progress: 100.0%
=====
PROCESSING IMAGES
=====
Progress: 0.0%
Progress: 7.69%
Progress: 15.38%
Progress: 23.08%
Progress: 30.77%
Progress: 38.46%
```

- Wynik:

Zdjęcie tekstu zostało przetworzone na tekst w pliku .txt

Zapisany w folderze [txt] w pliku (w tym przypadku 1.txt)

My Father-in-law is Lu Bu Chapter 199
Chapter 199 Escape (2)
Translated by Gamer

Two forces were running in the wilderness. A smaller group was at the front. A larger group was chasing them from the back. These were the Wolf Cavalry and the Yuzhou Cavalry.

Both forces were running with all their effort. The Wolf Cavalry at the front was running with all their effort so that they would not be torn into pieces by the Yuzhou Cavalry while the Yuzhou Cavalry was chasing with all their effort so that they could destroy the Wolf Cavalry. The general at the front of the Yuzhou Cavalry constantly try to make the speed of the cavalry move faster.

"Lu Bu! Run! Run! I want to see where do you want to run to!" Zhang Kai showed a ferocious face. He was impatient and wanted to rush into the Wolf Cavalry to tear them apart. He wants to kill all the wolf cubs with the blades of the Yuzhou Cavalry.

"Shoot the arrows! Shoot the arrows!" Zhang Kai shouted as he gave chase. They are about to catch up but a few horses out of reach. It would seem like this is as close as they could get. It was as if something tasty was put in

My Father-in-law is Lu Bu Chapter 199
Chapter 199 Escape (2)
Translated by Gamer

Two forces were running in the wilderness. A smaller group was at the front. A larger group was chasing them from the back. These were the Wolf Cavalry and the Yuzhou Cavalry.

Both forces were running with all their effort. The Wolf Cavalry at the front was running with all their effort so that they would not be torn into pieces by the Yuzhou Cavalry while the Yuzhou Cavalry was chasing with all their effort so that they could destroy the Wolf Cavalry. The general at the front of the Yuzhou Cavalry constantly try to make the speed of the cavalry move faster.

"Lu Bu! Run! Run! I want to see where do you want to run to!" Zhang Kai showed a ferocious face. He was impatient and wanted to rush into the Wolf Cavalry to tear them apart. He wants to kill all the wolf cubs with the blades of the Yuzhou Cavalry.