Dateien von PDF zu TXT konvertieren

Getestet mit `textract 1.5.0` und `python 2.7.13` .

In [ ]:

```
sourcepath = # Pfad zum Ordner, in dem die PDF-Dateien liegen, als String
targetpath = # Pfad zum Ordner, in den die TXT-Dateien gelegt werden sollen, als Sti
```

In [ ]:

```
import textract, os
```

In [ ]:

```
def convert(file_to_convert, file_to_save):
    """Receives a PDF file located at file_to_convert and creates from it a TXT file
    text = textract.process(file_to_convert)
    with open(file_to_save, "wb") as f:
        f.write(text)
```

In [ ]:

```
def convert_all(in_folder, out_folder):
    """Receives a collection of PDF files located in the in_folder
       and converts each file into a TXT file located in the out_folder.
    """
    for filename in os.listdir(in_folder):
        if filename.endswith(".pdf"):
            file_to_convert = "{}/{}".format(in_folder, filename)
            txtname = "{}.txt".format(filename[:-4])
            file_to_save = "{}/{}".format(out_folder, txtname)
            convert(file_to_convert, file_to_save)
```

Nun die eigentliche Konvertierung:

In [ ]:

```
convert_all(sourcepath, targetpath)
```

Ende.