
Tesseractwrap Documentation

Release 0.1.1

Greg Jurman, et al

Jun 05, 2017

Contents

1 Indices and tables	3
Python Module Index	5

Tesseractwrap is a ctypes/capi wrapper for [Tesseract OCR](#).

class `tesseractwrap.Tesseract` (*datadir=''*, *lang='eng'*)
Tesseract OCR object.

Parameters

- **datadir** – Tesseract data-directory with Tesseract training data.
- **lang** – The language of the image(s) to be OCR'd.

A simple example:

```
>>> from tesseractwrap import Tesseract
>>> from PIL import Image

>>> img = Image.open("test.png")
>>> tr = Tesseract()
>>> tr.ocr_image(img)
'The quick brown fox jumps ove\n\n'
```

clear ()

Clear the tesseract Image, and clean up any Tesseract run-data.

get_page_seg_mode ()

Returns the page analysis mode from Tesseract

get_rectangle ()

Get the bounding rectangle that tesseract is looking at inside of the image.

get_text ()

Get the text of the OCR'd image as a byte-string

get_utf8_text ()

Get the text of the OCR'd image as a string.

This function is kept for backwards compatibility with the 0.0 version of tesseractwrap.

ocr_image (*image*)

OCR an image returning the UTF8 text data.

Parameters *image* – image Image to be OCR'd by tesseract.

set_image (*image*)

Takes a PIL Image and loads it into Tesseract for further operations.

Note:: This function will automatically convert the image to Grayscale.

Parameters *image* – image Image to use in tesseract.

set_page_seg_mode (*mode=6*)

Set the page layout analysis mode.

Parameters *mode* – integer The page layout analysis mode. See PageSegMode class for options

set_rectangle (*left, top, width, height*)

Set the OCR detection bounding-box.

Parameters

- **left** – integer Pixels offset right from left of the image.
- **top** – integer Pixels offset down from the top of the image.
- **width** – integer Width of the bounding-box.

- **height** – integer Height of the bounding-box.

CHAPTER 1

Indices and tables

- [genindex](#)
- [modindex](#)
- [search](#)

t

tesseract, 1

C

`clear()` (tesseractwrap.Tesseract method), 1

G

`get_page_seg_mode()` (tesseractwrap.Tesseract method), 1

`get_rectangle()` (tesseractwrap.Tesseract method), 1

`get_text()` (tesseractwrap.Tesseract method), 1

`get_utf8_text()` (tesseractwrap.Tesseract method), 1

O

`ocr_image()` (tesseractwrap.Tesseract method), 1

S

`set_image()` (tesseractwrap.Tesseract method), 1

`set_page_seg_mode()` (tesseractwrap.Tesseract method), 1

`set_rectangle()` (tesseractwrap.Tesseract method), 1

T

Tesseract (class in tesseractwrap), 1

tesseractwrap (module), 1