

 This repository Search

Pull requests Issues Gist

 + 

dannnylo / **rtesseract**

Unwatch 9

★ Unstar 235

Fork 31

135 lines (97 sloc) 4.771 kB

Raw

Blame

History





rtesseract

gem version 1.3.0

build passing

coverage 100%

code climate 4.0

Ruby library for working with the Tesseract OCR.

REQUIREMENTS:

To work properly rtesseract are needed:

- Tesseract - Program
- ImageMagick - Program
- RMagick or mini_magick or quick_magick - Gem

Attention: Version 1.0.0 works fine with Ruby 2.0 and tesseract 3.0 and lower versions of rtesseract works fine with Ruby 1.8 and tesseract 2.0.4.

EXAMPLE USAGE

It's very simple to use rtesseract:

CONVERT IMAGE TO STRING

```
image = RTesseract.new("my_image.jpg")
image.to_s #Getting the value
```

CHANGE THE IMAGE

```
image = RTesseract.new("my_image.jpg")
image.source = "new_image.png"
image.to_s
```

TRANSFORM THE IMAGE

```
image = RTesseract.read("my_image.jpg") do |img|
  img = img.white_threshold(245)
  img = img.quantize(256, Magick::GRAYColorspace)
end
image.to_s
```

CONVERT PARTS OF IMAGE TO STRING

```
mix_block = RTesseract::Mixed.new("test.jpg") do |image|
  image.area(28, 19, 25, 25)
  image.area(180, 22, 20, 28)
```

```

    image.area(218, 22, 24, 28)
    image.area(248, 24, 22, 22)
  end
  mix_block.to_s

OR

mix_block = RTesseract::Mixed.new("test.jpg",{areas => [
  {:x => 28, :y=>19, :width=>25, :height=>25 },
  {:x => 180, :y=>22, :width=>20, :height=>28},
  {:x => 218, :y=>22, :width=>24, :height=>28},
  {:x => 248, :y=>24, :width=>22, :height=>22}
]})
mix_block.to_s

```

OPTIONS

Processors Options (*Rmagick* is default)

```

RTesseract.new("test.jpg", :processor => "mini_magick")
RTesseract.new("test.jpg", :processor => "quick_magick")

```

Note: For non process the image use NoneProcessor

```

RTesseract.new("test.jpg", :processor => "none")

```

Language Options

```

RTesseract.new("test.jpg", :lang => "deu")
* eng  - English
* deu  - German
* deu-f - German fraktur
* fra  - French
* ita  - Italian
* nld  - Dutch
* por  - Portuguese
* spa  - Spanish
* vie  - Vietnamese
Note: Make sure you have installed the language to tesseract

```

Other Options

```

RTesseract.new("test.jpg", options: :digits) # Only digit recognition

OR

RTesseract.new("test.jpg", options: [:digits, :quiet])

```

BOUNDING BOX: TO GET WORDS WITH THEIR POSITIONS

```

RTesseract::Box.new('test_words.png').words
# => [
#   {:word => 'If', :x_start=>52, :y_start=>13, :x_end=>63, :y_end=>27},
#   {:word => 'you', :x_start=>69, :y_start=>17, :x_end=>100, :y_end=>31},
#   {:word => 'are', :x_start=>108, :y_start=>17, :x_end=>136, :y_end=>27},
#   {:word => 'a', :x_start=>143, :y_start=>17, :x_end=>151, :y_end=>27},
#   {:word => 'friend,', :x_start=>158, :y_start=>13, :x_end=>214, :y_end=>29},
#   {:word => 'you', :x_start=>51, :y_start=>39, :x_end=>82, :y_end=>53},
#   {:word => 'speak', :x_start=>90, :y_start=>35, :x_end=>140, :y_end=>53},
#   {:word => 'the', :x_start=>146, :y_start=>35, :x_end=>174, :y_end=>49},
#   {:word => 'password,', :x_start=>182, :y_start=>35, :x_end=>267, :y_end=>53},
#   {:word => 'and', :x_start=>51, :y_start=>57, :x_end=>81, :y_end=>71},
#   {:word => 'the', :x_start=>89, :y_start=>57, :x_end=>117, :y_end=>71},
#   {:word => 'doors', :x_start=>124, :y_start=>57, :x_end=>172, :y_end=>71},
#   {:word => 'will', :x_start=>180, :y_start=>57, :x_end=>208, :y_end=>71},

```

```
#  {:word => 'open.', :x_start=>216, :y_start=>61, :x_end=>263, :y_end=>75}  
# ]
```

Contributing to rtesseract

- Check out the latest master to make sure the feature hasn't been implemented or the bug hasn't been fixed yet.
- Check out the issue tracker to make sure someone already hasn't requested it and/or contributed it.
- Fork the project.
- Start a feature/bugfix branch.
- Commit and push until you are happy with your contribution.
- Make sure to add tests for it. This is important so I don't break it in a future version unintentionally.
- Please try not to mess with the Rakefile, version, or history. If you want to have your own version, or is otherwise necessary, that is fine, but please isolate to its own commit so I can cherry-pick around it.

Links

- Github - github.com/dannnylo/rtesseract
- Rubygems - rubygems.org/gems/rtesseract

Copyright

Copyright © 2014 Danilo Jeremias da Silva. See LICENSE.txt for further details.

