PDF Analytics Client API Documentation

Release 1.0.5

PDF Analytics

Dec 11, 2017

Contents

1 Introduction **3**
- 1.1 Purpose . . . 3
- 1.2 Examples . . . 3
- 1.3 Dependencies . . . 3
- 1.4 Examples . . . 4

2 Installation **5**
- 2.1 Python 3 . . . 5
- 2.2 Python 2 . . . 5

3 References **7**
- 3.1 PDF Analytics Client . . . 7

4 Changelog **11**
- 4.1 1.0.5, December 11, 2017 . . . 11
- 4.2 1.0.4, November 25, 2017 . . . 11
- 4.3 1.0.3, November 24, 2017 . . . 11
- 4.4 1.0.2 . . . 11
- 4.5 1.0.1 . . . 11
- 4.6 1.0.0 . . . 12

5 Indices and tables **13**

Python Module Index **15**

PDFAnalytics is a web service which lets you use to verify PDF content for free.

This **python-pdf-analytics-client** library allows you to automate most common **PDFAnalytics** operations using Python 2 or Python 3.

python-pdf-analytics-client can be installed from the `pip` tool or downloaded from PyPI: https://pypi.python.org/pypi/python-pdf-analytics-client

The source is available on: https://github.com/pdf-analytics/python-pdf-analytics-client

Contents:

CHAPTER 1

Introduction

1.1 Purpose

The purpose of python-pdf-analytics-client is to provide a library that would help you to automate most common **PDFAnalytics** using its REST API.

python-pdf-analytics-client can verify :

- *textural content*, like text, font style, its location (using coordinates and page number)
- *image content* based on a locally stored image (pixel-by-pixel comparison), its actual size and location in the PDF
- *pdf-to-pdf* comparison, compare an uploaded PDF with a local one pixel-by-pixel, page-by-page

1.2 Examples

This example asserts there is the *figure.png* image on page 4 inside the *demo.pdf* PDF file.

```
>>> from pdf_analytics_client import APIClient
>>> server = APIClient(token='my_token')
>>> pdf_job = server.create_job(local_file='/Users/tester/demo.pdf')
>>> pdf_job.verify_image(local_img='/Users/tester/figure.png', top=24, left=64,␣
↪page=4)
```

1.3 Dependencies

python-pdf-analytics-client has only one dependency python-requests . All the dependencies shall be installed automatically when you will install the **python-pdf-analytics-client** module with pip.

1.4 Examples

You may find the examples at the GitHub repository : https://github.com/pdf-analytics/python-pdf-analytics-client/tree/master/examples

To run the **examples** you need to have registered to the site pdf-analytics and to get your token number.

To run the examples:

```
$ cd examples
# Install the dependecies
$ pip install -r requirements_examples
# Run the examples
$ behave -D token=<your_token_id>
```

CHAPTER 2

Installation

2.1 Python 3

To install python-pdf-analytics-client, install the `python-pdf-analytics-client` package from PyPI and dependencies.

On Windows, this is:

```
C:\Python36\pip.exe install python-pdf-analytics-client
```

(Though you may have a different version of Python installed other than 3.6)

On OS X, this is:

```
pip3 install python-pdf-analytics-client
```

On Linux, this is:

```
pip install python-pdf-analytics-client
```

2.2 Python 2

To install python-pdf-analytics-client, install the `python-pdf-analytics-client` package from PyPI and dependencies.

On Windows, this is:

```
C:\Python27\pip.exe install python-pdf-analytics-client
```

(Though you may have a different version of Python installed other than 2.7)

On OS X, this is:

```
pip install python-pdf-analytics-client
```

On Linux, this is:

```
pip install python-pdf-analytics-client
```

python-pdf-analytics-client will try to install the only dependency i.e. the python-requests library. This happens when pip installs python-pdf-analytics-client.

CHAPTER 3

References

This is a quickstart reference to using PyPDFAnalyticsClient.

3.1 PDF Analytics Client

The PDF Analytics Client is a high level module that enables the verification of the images and text of a local PDF file.

class `api_client`.**`APIClient`**(*token*, *url=u'https://pdf-analytics.com/api/'*)
Main API client class

`create_job`(*local_file*, *wait_to_complete=True*)
Create a PDF analysis job

Parameters

- **`local_file`** – the path of the local PDF file that needs to be uploaded to the server for the analysis
- **`wait_to_complete`** – wait for the PDF analysis to complete. Default value is True.

Returns The JobClass object,

`get_account_details`()
Get my account details

Returns a dictionary object with the user's account details { 'max_pdf_size_mb': 3, 'daily_max_count': 10, 'today_remaining': 4, }

`get_job`(*job_id*)
Get PDF analysis job

Parameters **`job_id`** – the PDF analysis job ID

Returns The JobClass object,

class api_client.**JobClass**(*id*, *client*)
Basic PDF analysis Job class

get_item(*left*, *top*, *page*, *type=u'any'*)
Get any item from the PDF (TODO: get figure)

Parameters

- **left** – Distance from the *left* of the page in *points*. Accepts single integer. e.g. 150
- **top** – Distance from the *top* of the page in *points*. Accepts single integer. e.g 200
- **page** – Number of page, e.g. 4
- **type** – Type of the the item.

Returns A JSON object with the item's information

get_metadata()
Get the metadata of the PDF

Returns A JSON object with the metadata of the PDF

get_status()
Get the status of the PDF analysis

Returns The analysis status as string. The string can be "In progress", "Error" or "Complete"

Return type str

verify_image(*path*, *left*, *top*, *page*, *compare_method=u'pbp'*, *tolerance=0.0*)
Verify a local image file exists in the PDF

Parameters

- **path** – The absolute or relative path of the locally stored image e.g. '/User/tester/apple.png'
- **left** – Distance from the *left* of the page in *points*. Accepts single integer. e.g. 150
- **top** – Distance from the *top* of the page in *points*. Accepts single integer. e.g 200
- **page** – Number of page, e.g. an integer 4 or a string 'all', 'last', '1-4'
- **compare_method** – Image comparison method
- **tolerance** – Comparison tolerance. Default value 0.0. Example: 0.02

Returns If the request is successful it returns 200. If it is not successful it returns the error message.

Return type JSON

verify_pdf(*path*, *excluded_areas=u''*, *tolerance=0.0*)
Verify a local PDF file with the uploaded job's PDF

Parameters

- **path** – The absolute or relative path of the locally stored PDF ilfe e.g. '/User/tester/report.pdf'
- **excluded_areas** – Excluded areas. List field. Example : [{'left':146, 'top':452, 'width':97, 'height':13,'page':2}, {'left': 414, 'top': 747, 'width': 45, 'height': 16, 'page': 'all'},]
- **tolerance** – Comparison tolerance. Default value 0.0. Example: 0.02

Returns If the request is successful it returns 200. If it is not successful it returns the error message.

Return type JSON

verify_text(*text*, *left*, *top*, *page*, *method=u'contains'*)
Verify a text exists in the PDF

Parameters

- **text** – The expected textural content. Accepts string. e.g. 'This is the expected text'
- **left** – Distance from the *left* of the page in *points*. Accepts single integer. e.g. 150
- **top** – Distance from the *top* of the page in *points*. Accepts single integer. e.g 200
- **page** – Number of page, e.g. an integer 4 or a string 'all', 'last', '1-4'
- **method** – Text comparison method

Returns If the request is successful it returns 200. If it is not successful it returns the error message.

wait_analysis_to_complete()
Wait for the PDF analysis to complete

After you submit the PDF to PDF Analytics website, the takes some seconds until it is ready to be used for verification.

Returns If the analysis is completed and returns *True* else if in 20 seconds the job is not complete, returns *False*

Return type bool

CHAPTER 4

Changelog

This document will track major changes in the project.

4.1 1.0.5, December 11, 2017

- Fix ModuleNotFoundError

4.2 1.0.4, November 25, 2017

- Fix the documentation
- Add logo to the documentation

4.3 1.0.3, November 24, 2017

- Add Python 3 support

4.4 1.0.2

- Add Categories and keywords in pip / setup

4.5 1.0.1

- Fix the SSL cert verification
- Fix the documentation

- Cosmetic changes to the python-behave examples

4.6 1.0.0

- First release

CHAPTER 5

Indices and tables

- genindex
- modindex
- search

Python Module Index

a

api_client, 7

Index

A

api_client (module), 7
APIClient (class in api_client), 7

C

create_job() (api_client.APIClient method), 7

G

get_account_details() (api_client.APIClient method), 7
get_item() (api_client.JobClass method), 8
get_job() (api_client.APIClient method), 7
get_metadata() (api_client.JobClass method), 8
get_status() (api_client.JobClass method), 8

J

JobClass (class in api_client), 7

V

verify_image() (api_client.JobClass method), 8
verify_pdf() (api_client.JobClass method), 8
verify_text() (api_client.JobClass method), 9

W

wait_analysis_to_complete() (api_client.JobClass method), 9