
odbrasil Documentation

Release 0.1

Christian S. Perone

June 30, 2012

CONTENTS

Release v0.1. (*Installation*)

odbrasil is an *Apache 2.0 licensed* Python module to extract Brazilian government open data. The aim of the project is to provide an unified, organized and well-documented API to extract and parse (typically into [Pandas](#) data structures) the government open data.

Today we have some projects doing scraping of the open data, but these projects doesn't offer a parse for [Pandas](#) and do not have an unified and organized API, most of them are just *scripts* created in a hurry on Hackatons and do not have any documentation.

The API we're working on is simple and easy-to-use, intended not only for programmers but also for statisticians that doesn't have a strong background development.

We **really need** the community support in order to cover a great part of the API available for the government open data, if you want to help, join us on Github.

I have chosen the [Pandas](#) because it is becoming the *lingua franca* of the Python data analysis toolkits and because it is integrated with [matplotlib](#) and [scipy/numpy](#) ecosystem.

Here is an example of what the API can do:

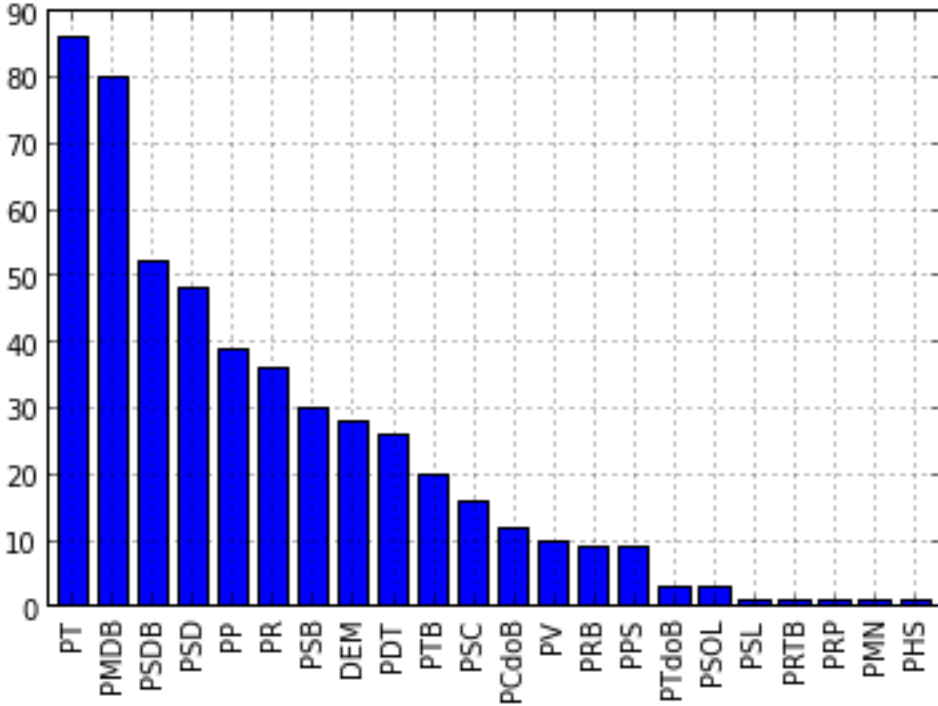
```
>>> from odbrasil.legislativo import camara
>>> api = camara.Deputados()
>>> deputados = api.get_deputados()
>>> deputados
<class 'pandas.core.frame.DataFrame'>
Int64Index: 512 entries, 0 to 511
Data columns:
anexo          512  non-null values
email          512  non-null values
fone           512  non-null values
gabinete       512  non-null values
idParlamentar 512  non-null values
nome           512  non-null values
nomeParlamentar 512  non-null values
partido        512  non-null values
sexo           512  non-null values
uf             512  non-null values
dtypes: object(10)

>>> vcounts = deputados.partido.value_counts()
>>> vcounts
PT          86
PMDB        80
PSDB        52
PSD         48
PP          39
PR          36
PSB         30
DEM         28
PDT         26
PTB         20
PSC         16
PCdoB       12
PV          10
PRB         9
PPS         9
PTdoB       3
PSOL        3
PSL         1
```

```

PRTB      1
PRP       1
PMN       1
PHS       1
>>> vcounts.plot(kind='bar')

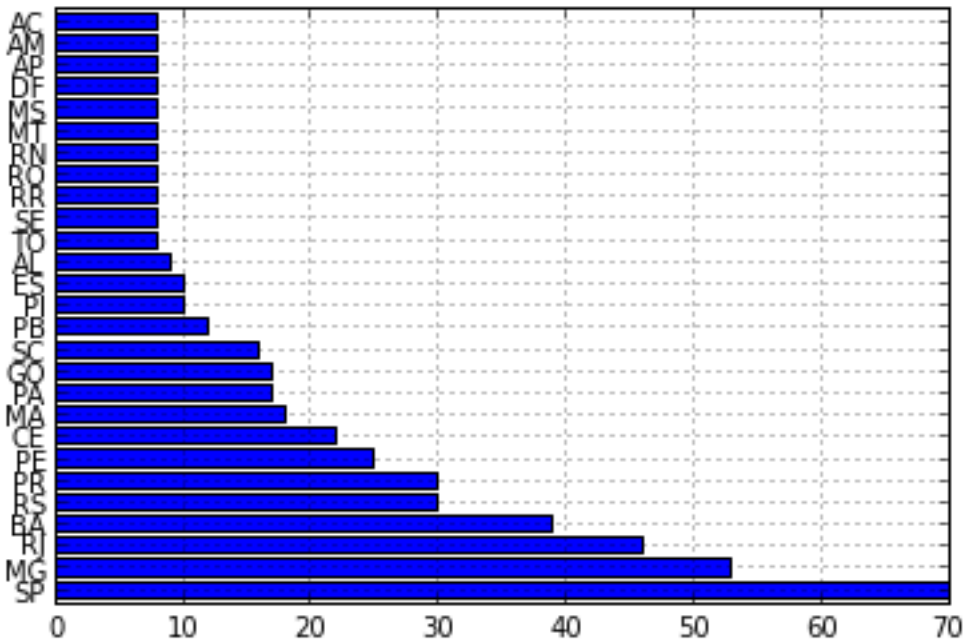
```



```

>>> uf_deputados = deputados.uf.value_counts()
>>> uf_deputados.plot(kind='barh')

```



See the [API documentation](#) for more information.

INSTALLATION

You can use **pip** to install **odbrasil** module and its dependencies, it is recommended that you have already installed **scipy/numpy** and **matplotlib** from your distro, in Ubuntu for instance:

```
sudo apt-get install python-numpy python-scipy python-matplotlib
```

And to install **odbrasil**:

```
pip install odbrasil
```

Simple and easy as that.

API DOCUMENTATION

If you are looking for information on a specific function, class or method, this part of the documentation is for you.

2.1 API

This part of the documentation covers all the interfaces of **odbrasil**.

Note: **odbrasil** usually returns the data extracted as a Pandas `DataFrame`, to get a better introduction on how to use it, see [Pandas documentation](#).

2.1.1 Legislativo

odbrasil.legislativo.camara

This module implements the methods to extract the information present on the government site for open data: <http://www2.camara.gov.br/transparencia/dados-abertos>

class `odbrasil.legislativo.camara.Deputados`

Bases: `odbrasil.legislativo.camara.RESTServiceClient`

This class is responsible by accessing, extracting and parsing the data from the `Deputados` government endpoint.

get_deputados (*format='pandas', **params*)

This method will get a `Deputados` list in various formats, use the `format` parameter to define which parameter you want to parse the data.

Parameters

- **format** – “pandas” or “xml”
- **params** – extra parameters will be redirected to Requests

Return type the parsed xml or the pandas `DataFrame`.

class `odbrasil.legislativo.camara.Orgaos`

Bases: `odbrasil.legislativo.camara.RESTServiceClient`

This class is responsible by accessing, extracting and parsing the data from the [Orgaos](#) government endpoint.

get_organos (*format='pandas', **params*)

This method will get a Orgaos list in various formats, use the `format` parameter to define which parameter you want to parse the data.

Parameters

- **format** – “pandas” or “xml”
- **params** – extra parameters will be redirected to Requests

Return type the parsed xml or the pandas *DataFrame*.

class `odbrasil.legislativo.camara.RESTServiceClient`

Bases: `object`

The base class used by other subclasses to retrieve data from the government webservices. If you want to subclass this class, you have to define two class variables on your subclass, called `base_url` and the expected `content_type`. See [Deputados](#) for reference.

This class is responsible for keeping the common functionality used by the service clients, like using the **User-Agent** as **odbrasil/1.0** for instance.

Note: you shouldn't use this `RESTServiceClient` on your application except if you really need to customize the internals of the REST client.

get (*service, **params*)

This method uses the `baseurl` parameter and concats the `service` parameter into it to create the request URL. Any extra param passed to this method by the `params` parameter will be redirected to the Requests request.

Parameters

- **service** – the service, i.e. ‘ObterDeputados’
- **params** – extra parameters to be used by Requests

Return type the Requests request response

`odbrasil.legislativo.camara.pandas_parse_deputados` (*xml_deputado_list*)

Method used to parse a xml parsed list of `deputado` elements into a pandas *DataFrame*.

Parameters `xml_deputado_list` – the xml parsed data returned by calling `Deputados.get_deputados()` with the `format` as ‘xml’ instead of ‘pandas’.

`odbrasil.legislativo.camara.pandas_parse_orgao` (*xml_orgao_list*)

Method used to parse a xml parsed list of `orgao` elements into a pandas *DataFrame*.

Parameters `xml_orgao_list` – the xml parsed data returned by calling `Orgaos.get_organos()` with the `format` as ‘xml’ instead of ‘pandas’.

LICENSE

Copyright 2012 Christian S. Perone

Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

3.1 Contributors

Christian S. Perone [[twitter](#)] [[blog](#)] [[github](#)].

PYTHON MODULE INDEX

O

`odbrasil.legislativo.camara, ??`