page Documentation

Release 0.1.5

Andre Rendeiro

May 28, 2021

CONTENTS

page is a Python library to perform Parametric Analysis of Gene Set Enrichment.

Head to the *Installation* to see installation instructions, to *Usage* for quick use, or have a look at the catalogue of available functions in the *API*.

CONTENTS

CONTENTS

1.1 Installation

1.1.1 With pip

`page` is available for Python 3 only. Is is tested in Python 3.6 and 3.7.

To install, simply do:

```
pip install page-enrichment
```

you might need to add a `--user` flag if not root or running in a virtual environment.

This will install all the Python dependencies needed too. See here a list of all Python dependencies used.

To install the latest development version:

```
pip install git+https://github.com/afrendeiro/page-enrichment.git#egg=page
```

1.2 Quick usage

1.2.1 Using the command-line interface (CLI)

`page` has a CLI for easy usage:

```
$ page-get-test-data test.csv
$ page -g KEGG_2016 test.csv output.csv
```

1.2.2 Interactive usage through the API

Example usage of the *page.page()* function:

```python
from page import page, get_test_data

vect = get_test_data()
results = page(vect, gene_set_libraries=['KEGG_2016'])
```

The `page.get_test_data()` function serves simply the purpose of getting a vector of fold-changes to use with PAGE.

1.3 Configuration and logging

1.3.1 Gene set libraries and configuration

`page` uses gene sets from Enrichr gene set libraries.

Enrichment of a few is performed by default but this can be customizable by the user.

The user can customize `page` in two ways:

- In a YAML file located in `$HOME/.page.config.yaml`;
- At run time with through the `gene_set_libraries` option in the *page.page()* function.
- At run time with the `-g` or `--gene-set-libraries` option if using the command-line interface.

The YAML configuration file is entirely optional, but does allow more flexibility and ease specially when selecting a large number of gene set libraries (each with its own gene sets).

To see which gene set libraries are available, go to https://amp.pharm.mssm.edu/Enrichr/#stats.

`page` does not need any static files with genes, but gets automatically what is needed and keeps a cache of what has been used in order to save API calls and greatly improve speed.

1.3.2 Logging

`page` will log its operations and errors using the Python standard logging library.

This will happen by default to standard output (sys.stdout) but also to a file in `$HOME/.page.log.txt`.

The location of the log file and the level of events to be reported can be customized in the `page.setup_logger()` function.

1.4 API

1.4.1 page

Parametric Analysis of Gene Set Enrichment (PAGE).

> **Parameters**
>
> - **parameter_vector** (`Series`) – Pandas series with index as genes and values as 'parameters' (e.g. fold-changes, expression values)
> - **gene_set_libraries** (`list`, optional) – Gene set libraries to use. Defaults to values in default configuration file or in `~/.page.config.yaml`. To see all available values go to https://amp.pharm.mssm.edu/Enrichr/#stats
>
> **returns** Data frame with enrichment statistics for each term in each gene set library.
>
> **rtype** `DataFrame`

1.5 Testing

`page` does not have many dependencies, but to make sure everything is correctly configured, the user is encouraged to test the library prior to use.

In order to do this, install testing requirements and simply run `pytest`:

```
pip install page-enrichment[testing]
pytest --pyargs page
```

Pytest will output summary results (see for example) and further outputs can be seen in `${TMPDIR}/pytest-of-${USER}/` or `/tmp/pytest-of-${USER}/` if $TMPDIR is not defined.

1.6 Changelog

All notable changes to this project will be documented in this file.

The format is based on Keep a Changelog and this project adheres to Semantic Versioning.

1.6.1 [0.1.0] - 2020-02-06

Added

- Basic functionality to perform enrichmemnt, get gene sets from Enrichr and CLI

1.7 Indices and tables

- genindex
- modindex
- search

TWO

LINKS

- Documentation: http://toolkit.readthedocs.io/

- Issues and source code: https://github.com/afrendeiro/toolkit

PYTHON MODULE INDEX

p

M

module
    page.page, 4

P

page.page
    module, 4