

---

# **snowballing Documentation**

***Release 0.1.8***

**Joao Felipe Pimentel**

**Jul 28, 2017**



---

## Contents:

---

<b>1</b>	<b>Contact</b>	<b>3</b>
1.1	Getting started	3
1.2	approaches.py	6
1.2.1	Classes	6
1.2.1.1	Group	6
1.2.1.2	GroupUnrelated	6
1.2.1.3	Item	6
1.2.2	Functions	6
1.2.2.1	name	6
1.2.2.2	get_approaches	6
1.2.2.3	wcite	6
1.2.2.4	wlatex_name	6
1.2.2.5	wcitea	6
1.3	common_places.py	6
1.3.1	Functions	6
1.3.1.1	conference	6
1.3.1.2	journal	6
1.3.1.3	magazine	6
1.4	config.py	6
1.5	dbindex.py	6
1.5.1	Functions	6
1.5.1.1	citation_file	6
1.5.1.2	year_file	6
1.5.1.3	places_file	6
1.5.1.4	this_file	6
1.5.1.5	discover_year	6
1.5.1.6	increment_char	6
1.5.1.7	increment_str	6
1.5.1.8	parse_varname	6
1.6	dbmanager.py	6
1.6.1	Relevant functions	6
1.6.1.1	insert	6
1.6.1.2	set_attribute	6
1.6.1.3	rename_work	6
1.6.1.4	insert_work	6
1.6.1.5	insert_citation	6

1.6.1.6	remove_target_citation	6
1.6.1.7	remove_source_citation	6
1.6.2	Other Functions	6
1.6.2.1	rename_lines	6
1.6.2.2	rename_citation	6
1.6.2.3	citation_operation	6
1.6.2.4	work_operation	6
1.6.2.5	save_lines	6
1.6.2.6	read_file	6
1.6.2.7	is_assign_to_name	6
1.6.2.8	is_call_statement	6
1.6.3	Operation classes	6
1.6.3.1	ReplaceOperation	6
1.6.3.2	DelOperation	6
1.6.3.3	AddKeywordOperation	6
1.6.3.4	InsertOperation	6
1.6.3.5	DetectOperation	6
1.6.4	Visitor classes	6
1.6.4.1	EditVisitor	6
1.6.4.2	BodyVisitor	6
1.6.4.3	CitationVisitor	6
1.7	graph.py	6
1.7.1	Classes	6
1.7.1.1	Graph	6
1.7.1.2	GraphConfig	6
1.7.2	Functions	6
1.7.2.1	create_graph	6
1.7.2.2	set_positions	6
1.7.2.3	getcolors	6
1.7.2.4	getsvgcolors	6
1.8	jupyter_utils.py	6
1.8.1	Functions	6
1.8.1.1	display_cell	6
1.8.1.2	idisplay	6
1.8.1.3	find_line	6
1.8.1.4	invoke_editor	6
1.8.1.5	new_button	6
1.8.1.6	work_button	6
1.9	models.py	6
1.9.1	Classes	6
1.9.1.1	Work	6
1.9.1.2	Site	6
1.9.1.3	Email	6
1.9.1.4	Citation	6
1.9.1.5	Year	6
1.9.1.6	Database	6
1.10	operations.py	6
1.10.1	Database	6
1.10.1.1	reload	6
1.10.1.2	work_by_varname	6
1.10.1.3	find	6
1.10.1.4	load_work_map_all_years	6
1.10.1.5	load_work	6
1.10.1.6	load_citations	6

1.10.1.7	load_places_vars . . . . .	6
1.10.1.8	load_work_map . . . . .	6
1.10.1.9	citation_text . . . . .	6
1.10.1.10	find_citation . . . . .	6
1.10.1.11	find_global_local_citation . . . . .	6
1.10.2	BibTeX . . . . .	6
1.10.2.1	work_to_bibtex_entry . . . . .	6
1.10.2.2	work_to_bibtex . . . . .	6
1.10.2.3	match_bibtex_to_work . . . . .	6
1.10.2.4	compare_paper_to_work . . . . .	6
1.10.2.5	find_work_by_info . . . . .	6
1.10.2.6	bibtex_to_info . . . . .	6
1.10.2.7	set_display . . . . .	6
1.10.2.8	set_pyref . . . . .	6
1.10.2.9	set_place . . . . .	6
1.10.2.10	extract_info . . . . .	6
1.10.2.11	info_to_code . . . . .	6
1.11	scholar.py . . . . .	6
1.12	selenium_scholar.py . . . . .	6
1.13	snowballing.py . . . . .	6
1.13.1	Widgets . . . . .	6
1.13.1.1	Converter . . . . .	6
1.13.1.2	ArticleNavigator . . . . .	6
1.13.1.3	BackwardSnowballing . . . . .	6
1.13.1.4	ForwardSnowballing . . . . .	6
1.13.1.5	ScholarUpdate . . . . .	6
1.13.2	Analysis . . . . .	6
1.13.2.1	create_provenance . . . . .	6
1.13.2.2	log_to_provn . . . . .	6
1.13.2.3	snowballing . . . . .	6
1.13.2.4	Step . . . . .	6
1.14	utils.py . . . . .	6
1.14.1	Functions . . . . .	6
1.14.1.1	match_any . . . . .	6
1.14.1.2	compare_str . . . . .	6
1.14.1.3	setitem . . . . .	6
1.14.1.4	consume . . . . .	6
1.14.1.5	import_submodules . . . . .	6
1.14.1.6	adjust_point . . . . .	6
1.14.1.7	multiline_wrap . . . . .	6
1.14.1.8	lines_len_in_circle . . . . .	6
1.14.1.9	text_y . . . . .	6
1.14.2	Classes . . . . .	6
1.14.2.1	MultiLine . . . . .	6
1.14.2.2	Point . . . . .	6
1.15	Contributing . . . . .	6
1.15.1	Notes . . . . .	7
1.16	License . . . . .	7



This project provides tools for performing a literature review through snowballing. It includes Jupyter Notebook widgets that assist the backward and forward steps of literature snowballing, notebooks that assist in inserting citations in the database, and notebooks for analyzing the snowballing and producing citation graphs, publication place histograms, and a summarization of the snowballing steps.

This package was tested on Python 3.6 using Windows, but it should support Python > 3.5 in any operating system. Please, open an issue if it is not the case.





Do not hesitate to contact me:

- João Felipe Pimentel [joaofelipenp@gmail.com](mailto:joaofelipenp@gmail.com)

## Getting started

To install the tool, you should follow these instructions:

- First, install the python package:
- `$ pip install snowballing`
- Then, download and install the latest [geckodriver](#)
- In my case, I put the 'geckodriver.exe' file in the Firefox directory (C:\Program Files (x86)\Mozilla Firefox)
- And I add the directory to the PATH environment variable
- If you want to export the snowballing history (provenance), you must also install the following tools
- [provtoolbox](#)
- [GraphViz](#)

For starting a new literature review project, please run:

```
$ snowballing start literature
```

This command will create a directory called `literature` (you are free to use other name in the command) with the notebooks for performing the snowballing and analyzing it, and an example database.

Inside the directory, start Jupyter:

```
$ cd literature
$ jupyter notebook
```

And open the file `Index.ipynb`. This file contains all the instructions for understanding the database and performing the snowballing.



## approaches.py

### Classes

Group

GroupUnrelated

Item

### Functions

name

get\_approaches

wcite

wlatex\_name

wcitea

## common\_places.py

### Functions

conference

journal

magazine

## config.py

## dbindex.py

### Functions

citation\_file

year\_file

places\_file

this\_file

discover\_year

increment\_char

increment\_str

parse\_varname

..

For installing in development mode, clone the repository and use `pip install -e`:

```
$ git clone git@github.com:JoaoFelipe/snowballing.git
$ cd snowballing
$ pip install -e snowballing
```

## Notes

This project started as part of a literature snowballing. The tools were developed out of necessity in an ad hoc way. Thus, it has some bad design decisions, such as using Python scripts as a database, and choosing field names that are far from ideal.

Contributions to fix these and other issues are welcome! If you are going to change a field name that is used by the tool, please, try to make it configurable.

## License

The MIT License (MIT)

Copyright (c) 2017 Joao Felipe Pimentel <[joaofelipenp@gmail.com](mailto:joaofelipenp@gmail.com)>

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.