Read Me / Kick-start guide of the Project

# Installing Python and required dependencies via PyPi (PIP)

There is not particular restriction of which version of Python and/or its packages should be used.

However, to avoid version collision, you may consider using virtual environments (like from [venv](https://docs.python.org/3/library/venv.html), [Anaconda](https://docs.anaconda.com/anaconda/) or [Miniconda](https://docs.anaconda.com/miniconda/)).

**Example.** If you would like to run a script written for Python 2 and this script (which is written in Python 3), you should use a virtual environment.

All the packages required are listed in the “requirements.txt”. Run the following command to install the Python packages:

pip install -r requirements.txt

# Installing MongoDB

For optimal performance (especially in RSV read-write), MongoDB is used as the control centre and the central storage for all scrappers deployed.

A MongoDB server could be downloaded from the publisher’s website: <https://www.mongodb.com/try/download/community>

The scrappers and the database will connect over port 27017.

# Data Backup and Restore

A BSON data (including the RSVs and the metadata) is included in this archive.

Use “mongodbrestore” to import the BSON, and use “mongodbdump” to export the BSON.   
Both software should be installed on operators’ machine instead of the server.  
Both pieces of software are part of the “MongoDB Database Tools” and can be downloaded from the publisher’s website:   
<https://www.mongodb.com/docs/database-tools/installation/installation/>

# Package Structure

|  |  |
| --- | --- |
| ├── data  │   └── joined\_msvs.csv | The merged search volumes (MSVs) of all keywords. |
| ├── gtdownloader      ├── download.py  └── … | Scripts used to download RSVs using web scraping. |
| ├── scripts |  |
| │   ├── clustering  │   │   ├── convert\_and\_plot\_polar\_coordinates.py  │   │   ├── plot\_unscalted\_polar\_projection.py  │   │   ├── plot\_unscalted\_polar\_projection.v2.py  │   │   └── sample\_plot\_cosine\_wave.py | Scripts used to create the projections (unscaled or scaled) for MSVs and centroid calculations. |
| │   ├── graphics  │   │   ├── export\_to\_gephi.py  │   │   ├── plot\_centroid\_only.py  │   │   ├── plot\_msv\_after\_point.py  │   │   ├── plot\_msv\_after\_point\_with\_ma.py  │   │   ├── plot\_msv.experiment\_twiny.py  │   │   ├── plot\_msv\_longcovid.py  │   │   ├── plot\_msv\_per\_plot.py  │   │   ├── plot\_msv\_per\_plot.v2.py  │   │   ├── plot\_msv\_per\_plot.v3a.py  │   │   ├── plot\_msv\_per\_plot.v3b\_after.py  │   │   ├── plot\_msv\_per\_plot.v3b.py  │   │   ├── plot\_msv\_per\_plot.v3.full.py  │   │   ├── plot\_msv.py  │   │   ├── plot\_msv\_v2.py  │   │   ├── plot\_msv\_v3.py  │   │   ├── plot\_netstats.py  │   │   ├── plot\_rescaled\_polar\_ring\_v2\_per\_keywords.py  │   │   ├── plot\_rescaled\_polar\_ring\_v2.py  │   │   └── plot\_s\_t.py | Scripts used to plot graphics. |
| │   ├── info  │   │   ├── covid\_events.py  │   │   ├── keywords.py | The different definitions used in this project. |
| │   ├── msvs  │   │   ├── generate\_msv.py  │   │   ├── join\_msv.py  │   │   └── segement\_msvs.py | Scripts for creating MSV from multiple RSVs. |
| │   └── rsvs  │   ├── check\_incomplete\_rsv.py  │   ├── download\_rsv.py  │   ├── move\_incomplete\_rsv.py  │   └── read\_rsv.py | Scripts for inspecting the downloaded RSVs. |