

download_agora.py

The first function to be called in the main function is 'get_ty_links'.

get_ty_links() :-

Here we initialize two arrays list years = [] and year_links = [] using for loop we download the dataset for the needed period. We give the particular year in the year list and the corresponding link for that year will be appended in year_links list. Again we initialize two lists as tys = [] and ty_links = []

we get the dataset from " agora.ex.nii.ac.jp/digital-typhoon". This website contains all the records of Typhoons from 1951 to 2022. For example if we take a particular year all the cyclones which are occurred during that particular year is also be recorded. All the records for that particular year (2021 or 202123) will be stored in tys[] list. And the corresponding links will be stored in ty_links[] list. For extracting these links we will use BeautifulSoup library. This library is used for web scraping. Web scraping is the process of collecting data from the internet. soup.find_all() function is used for finding out all tags

with the specified tag name or id. The 'get-ty-links()' function will return tys[] and ty-links[] lists. Next we are calling 'download-pmgs(ty, links)' function which consists of two arguments tys and links which are returned from 'get-ty-links()'. The 'download-pmgs()' function is used for downloading the images for dataset from the website.

download-pmgs(tys, ty-links):

We are creating a folder to upload the dataset.

By using for loop we are downloading the images and also we get the information of wind and central pressure.

First the ty-links attribute is passed to the httpget() function.

This function will return the html code for that particular link. After getting the html page the content is parsed using the 'Beautifulsoup' method. In all the links (tag a) will be stored in a list named a-list. These links are extracted from that page using "find-all" method. In that link the wind,

central pressure values are present.

The cyclone image is downloaded using

"httpdownload" function. After downloading

the image, the image name is created.

The format is "YearMonth - Timestamp - wind -

central pressure".

httpget(url) :-

In this function the ~~se~~ requests.get() function is called. The argument is url and timeout. This function will return a response object.

This function returns the content in the url in html format.