WeatherComPdoc

View Source

class WeatherCom: View Source

WeatherCom is a class to extract the date from weather.com in HTML format and transform the data into a dictionary to store them in the mongodb database.

We have used multimultiple python library such as requests, BeautifulSoup, datetime,time and MongoClient

- requests: it is a liberary that is used to allows the code to send HTTP/1.1 requests in simple way.
- BeautifulSoup: this liberary is used to extract the data out from the html file
- datetime, time: both are used to manipulate dates and times
- MongoClient: it is a liberary which is used to write to MongoDB

WeatherCom() View Source

we use the function now() from datatime module to get current time and date and store it as a global variable in self.now

The URL has 6edbe6f43ea5f01b5ce04bde2bf64d098511d4950e718605916e408bf8766b7b which is return the forecast of Pittsburgh weather in Fehrenhite.

The data that we have extract from the URL is stored into the response variable -page-. This code line has been repeated three times with three different timestep: today, tenday, and hourbyhour forecast.

We have used BeautifulSoup to parse html content of the page object, so the data can easliy extracted from HTML. The prased HTML pags is stored in soup object.

The function importInformation(soup, timeStep) is called three times for three different timeStep, for each time we pass the soup and the timeStep as parameter to return the information that we have asked for in a list. All three list is a global variable. .

The last step that the function sleep has been used because one of the Weather.com restrictions it is not allowed to access the website within 10 second

def importInformation(self, soup, timeStep): View Source

The function importInformation(soup, timeStep) is called to extract the required attribute from the HTML tags. This function has three cases, for today information, tenday information and hourly information. Each case of those three cases has its own HTML tags, so the tag that we have extract the Max_temp (in F) for today time step is different than the one that is used to extract Max_temp (in F) of tenday time step. The result is returned in a list of dictionay where each dictionary has (day, Temp, Max_temp (in F), Min_temp (in F), precipitationProb, Description and Time collected). However, the hourly dictionary has one more attribute which is (hour).

def WCMongoWrite_tenday(self, URL, collectionname): View Source

WCMongoWrite_tenday is a function which is used to write the inforamtion of the global list which is tenday_list to the mongodb.

- client object is creating a connection to the mongodb.
- After that the varialbe db is accessing Database object by creating a new one or accessing an exciting database.
- Finally, tenday_collection objec is accessing the Collection that is used to store the data in by using insert_one() function.

def WCMongoWrite_hourly(self, URL, collectionname): View Source

WCMongoWrite_hourly is a function which is used to write the inforamtion of the global list which is hourly_list to the mongodb.

- client object is creating a connection to the mongodb.
- After that the varialbe db is accessing Database object by creating a new one or accessing an exciting database.
- Finally, hourly_collection objec is accessing the Collection that is used to store the data in by using insert_one() function.

def WCMongoWrite_today(self, URL, collectionname): View Source

WCMongoWrite_today is a function which is used to write the inforamtion of the global list which is today_list to the mongodb.

- client object is creating a connection to the mongodb.
- After that the varialbe db is accessing Database object by creating a new one or accessing an exciting database.
- Finally, today_collection object is accessing the Collection that is used to store the data in by using insert_one() function.