Main modules:

1. api = allows the user to manipulate with website including scraping in websites and search in the content
2. flight and article are entities in which later on use to parse the website content into json
3. page Crawler is responsible to scrap the website and download content
4. locator module = this module holds all CSS locators for each website to be scarp
5. save Content Module is responsible to parse web content into Json file

main function:

1. save function in saveContentModule, takes ant data\_list and json file name. and parse the data into json file
2. PageCrawler get content is an override function. each page overrides this function in order to download page content with the correct and exclusive locators
3. search function in search file is an override function. each main page overrides this function in order to parse web data into Json correctly

solution:

This program is a web scraping program.

the program uses selenium in order to open pages move between URLs and extract content

The program starts with extracting all content of requested URLS

Then saveing all raw content into json.files

After extraction the program gives query service from all content files.

Configures:

The program already holds chromedriver.exe with in the directory

Please install the following (if needed):

Pip install selenium, pip install webdriver-manager

David Alfia

0506823212

david.alfia@gmail.com