

```

1  import time
2
3  from urllib import parse
4  from selenium import webdriver
5  from webdriver_manager.chrome import ChromeDriverManager
6
7  from Logger import Logging
8
9  # For selenium driver implementation on heroku
10 chrome_options = webdriver.ChromeOptions()
11 chrome_options.add_argument('--disable-gpu')
12 chrome_options.add_argument('--no-sandbox')
13 chrome_options.add_argument("disable-dev-shm-usage")
14
15 logger_obj = Logging('Advance Image Downloader') # Creating a custom based logger
16 logger_obj.initialize_logger() # Instantiating the logger object
17
18
19 class ImageScrapperClass:
20
21     def __init__(self, no_images):
22         """
23         This function will instantiate the query email and no_images parameter
24         :param no_images: Number of images want
25         """
26         try:
27             self.no_images = no_images
28             self.browser = webdriver.Chrome(executable_path=ChromeDriverManager().install(),
29                                             chrome_options=chrome_options)
30         except Exception as e:
31             logger_obj.print_log('(Scrapper.py(__init__)) - Something went wrong ' + str(e), '
exception')
32             raise Exception(e)
33
34     def get_request(self, search_query):
35         """
36         This function will open the chrome browser for fetching operations
37         """
38         try:
39             # Parsing the search query for searching over google
40             search_query = parse.quote_plus(search_query)
41
42             url = 'https://www.google.com/search?hl=jp&q=' + search_query + '&btnG=Google+
Search&tbs=0&safe=off&tbm' \
43                 '=isch '
44             self.browser.get(url)
45         except Exception as e:
46             logger_obj.print_log('(Scrapper.py(get_request)) - Something went wrong ' + str(e), '
exception')
47             raise Exception(e)
48
49     def store_url(self, thumbnail_images, current_count, no_thumbnail_images, final_images,

```

```

49 req_id, email, cassandra):
50     """This function inserts the images URL inside the database and then returns the this
        URL in the set format
51     :param final_images: Images URL found
52     :param thumbnail_images: Thumbnails selected
53     :param current_count: Current count of urls found
54     :param no_thumbnail_images: Number of images selected
55     :param cassandra: Cassandra object
56     :param email: Email ID of user
57     :param req_id: Unique Request ID
58     :return: set of url found
59     """
60     try:
61         task_finished = False
62         for img in thumbnail_images[current_count: no_thumbnail_images]:
63             # Try to click on every thumbnail so that it can open at side
64             try:
65                 img.click()
66                 time.sleep(0.5)
67             except Exception as e:
68                 logger_obj.print_log('(Scrapper.py(store_url)) - Something went wrong ' + str(e)
69                                     ), 'info')
70                 continue
71             # Extract the image urls from the thumbnail which is opened
72             opened_images = self.browser.find_elements_by_css_selector('img.n3VNCb')
73             for actual_image in opened_images:
74                 if actual_image.get_attribute('src') and 'http' in actual_image.get_attribute('src'
75                                     ):
76                     if not actual_image.get_attribute('src') in final_images:
77                         final_images.add(actual_image.get_attribute('src'))
78                         cassandra.insert_url(req_id, email, actual_image.get_attribute('src'))
79                         current_count += 1
80             # Condition to not return the more images then what is expected from the user
81             if current_count >= self.no_images:
82                 task_finished = True
83                 break
84
85         return final_images, current_count, task_finished
86     except Exception as e:
87         logger_obj.print_log('(Scrapper.py(store_url)) - Something went wrong ' + str(e), '
exception')
88         raise Exception(e)
89
90     def fetch_thumbnails(self, req_id, email, cassandra):
91         """
92         This function will fetch the thumbnails of images from the Chrome
93         :param email: Email ID of the user
94         :param cassandra: Cassandra object
95         :param req_id : Unique request id of the request
96         """

```

```

97     try:
98         final_images = set()
99         current_count = 0
100        task_finished = False
101
102        while current_count < self.no_images and not task_finished:
103            # Scroll to the end of the result section
104            self.scroll_to_end()
105
106            # If show more button exists then click
107            if self.browser.find_element_by_css_selector('.mye4qd'):
108                self.browser.execute_script('document.querySelector(".mye4qd").click()')
109
110            # Get all the thumbnail result
111            thumbnail_images = self.browser.find_elements_by_css_selector('img.Q4LuWd')
112            no_thumbnail_images = len(thumbnail_images)
113
114            final_images, current_count, task_finished = self.store_url(thumbnail_images,
115                                current_count,
116                                no_thumbnail_images, final_images, req_id,
117                                email, cassandra)
118
119            logger_obj.print_log('Images URL has been downloaded', 'info')
120
121            # Writing all the links in to the files
122            print('The number of images are {}'.format(len(final_images)))
123
124            # Closing the browser connection
125            self.close_browser()
126
127        except Exception as e:
128            logger_obj.print_log('(Scrapper.py(fetch_thumbnails) - Something went wrong ' +
129                                str(e), 'exception')
130            raise Exception(e)
131
132        def scroll_to_end(self):
133            """
134            This function scrolls the browser window to the end of the document body
135            """
136            try:
137                self.browser.execute_script('window.scrollTo(0, document.body.scrollHeight);')
138                time.sleep(1.5)
139            except Exception as e:
140                logger_obj.print_log('(Scrapper.py(scroll_to_end) - Something went wrong ' + str(e)
141                                    ), 'exception')
142                raise Exception(e)
143
144        def close_browser(self):
145            """
146            This function closes the browser object
147            """
148            try:

```

```
146         self.browser.close()
147     except Exception as e:
148         logger_obj.print_log('(Scrapper.py(close_browser) - Something went wrong ' + str(e
149                                ), 'exception')
150         raise Exception(e)
```