

Nama : Raihan Rahmanda Junianto

NIM : 222112303

Kelas : 3SD2

Penugasan Praktikum 12 Information Retrieval

Praktikum kali ini memiliki tujuan untuk melakukan scraping ulasan pengguna yang diperoleh melalui Google Maps pada tempat yaitu Jakarta Aquarium dengan link: <https://maps.app.goo.gl/SyXMj8rCRsco4JM38>.

Berikut disajikan keseluruhan kode untuk melakukan tugas tersebut.

```
"""This script serves as an example on how to use Python
    & Playwright to scrape/extract data from Google Maps"""

from playwright.sync_api import sync_playwright
from dataclasses import dataclass, asdict, field
import pandas as pd
import argparse
import time
import sys

@dataclass
class Review:
    """holds Review data"""
    id_review: str = None
    name: str = None
    review_text: str = None

@dataclass
class ReviewList:
    """holds list of Review objects,
    and save to both excel and csv
    """
    review_list: list[Review] = field(default_factory=list)

    def dataframe(self):
        """transform Review_list to pandas dataframe

        Returns: pandas dataframe
        """
        return pd.json_normalize((asdict(review) for review in
self.review_list), sep="_")
```

```

def save_to_excel(self, filename):
    """saves pandas dataframe to excel (xlsx) file

    Args:
        filename (str): filename
    """
    self.dataframe().to_excel(f"{filename}.xlsx", index=False)

def save_to_csv(self, filename):
    """saves pandas dataframe to csv file

    Args:
        filename (str): filename
    """
    self.dataframe().to_csv(f"{filename}.csv", index=False)

def main():
    with sync_playwright() as p:
        browser = p.chromium.launch(headless=False)
        page = browser.new_page()
        page.goto("https://maps.app.goo.gl/SyXMj8rCRsco4JM38", timeout=60000)

        page.wait_for_timeout(5000)

        page.locator('button:has-text("Ulasan lainnya")').click();

        review_list = ReviewList()

        page.wait_for_timeout(5000)

        for i in range(1, total):
            review = Review()
            review_element =
page.query_selector('//*[ @id="QA0Szd"]/div/div/div[1]/div[2]/div/div[1]/div/div[2]/div[9]/div['+str(3*i+-2)+']/div/div/div[2]/div[2]/div[1]/button')
            review_id = review_element.get_attribute('data-review-id')
            reviewer_name_xpath =
'//*[ @id="QA0Szd"]/div/div/div[1]/div[2]/div/div[1]/div/div/div[2]/div[9]/div['
'+str(3*i+-2)+']/div/div/div[2]/div[2]/div[1]/button/div[1]'
            reviewer_name = page.locator(reviewer_name_xpath).inner_text()

            review_xpath = '//*[ @id="'+review_id+'"]'

            if "... Lainnya" in page.locator(review_xpath).inner_text():
                button_lainnya_xpath =
'//*[ @id="'+review_id+'"]/span[2]/button'
                page.locator(button_lainnya_xpath).click();

```

```

        review_text = page.locator(review_xpath).inner_text()
        review.id_review = review_id
        review.name = reviewer_name
        review.review_text = review_text

        review_list.review_list.append(review)

        page.mouse.wheel(0, 7000)
        page.wait_for_timeout(3000)
        # Print 14 huruf terakhir review id dan total review yang sudah di
scrape
        print(f"Review ID: ...{review_id[-14:]} | Currently Scraped: {i}",
end='\r')
        sys.stdout.flush()

        review_list.save_to_excel("D:/RAIHAN STIS/Perkuliahan/SEMESTER
5/Praktikum INFORMATION RETRIEVAL/Pertemuan
(12)/jakarta_aquarium_review_data")
        browser.close()

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument("-t", "--total", type=int)
    args = parser.parse_args()

    # total number of products to scrape. Default is 10
    if args.total:
        total = args.total + 1
    else:
        total = 20

    main()

```

Pada kode di atas menggunakan beberapa library serta modul baru dengan masing-masing kegunaan sebagai berikut.

- 'sync_playwright' dari modul 'playwright.sync_api' berisi modul Playwright yang menyediakan antarmuka sinkron untuk otomatisasi browser.
- 'dataclass, asdict, field' dari modul 'dataclasses' digunakan untuk membuat kelas data (data class). Dalam konteks ini ditunjukkan oleh kelas ReviewList.
- 'argparse' digunakan untuk mengolah argumen baris perintah. Dalam konteks ini digunakan untuk mengatur argument pada fungsi main.
- 'time' digunakan untuk menangani waktu atau penundaan, seperti timeout, dan lain sebagainya.

- 'sys' merupakan modul sistem untuk interaksi dengan interpreter Python. Dalam hal ini, digunakan untuk menangani buffer output menggunakan perintah stdout.

```
from playwright.sync_api import sync_playwright
from dataclasses import dataclass, asdict, field
import pandas as pd
import argparse
import time
import sys
```

Selain itu, terdapat juga fungsi 'dataframe(self)' yang digunakan untuk mengubah Review_list menjadi sebuah dataframe berbasis pandas serta fungsi 'save_to_excel' dan 'save_to_csv' untuk menyimpan output scraping ke dalam dua format file, yaitu .xlsx dan .csv.

```
def dataframe(self):
    """transform Review_list to pandas dataframe

    Returns: pandas dataframe
    """
    return pd.json_normalize([asdict(review) for review in self.review_list], sep="_")

def save_to_excel(self, filename):
    """saves pandas dataframe to excel (xlsx) file

    Args:
        filename (str): filename
    """
    self.dataframe().to_excel(f"{filename}.xlsx", index=False)

def save_to_csv(self, filename):
    """saves pandas dataframe to csv file

    Args:
        filename (str): filename
    """
    self.dataframe().to_csv(f"{filename}.csv", index=False)
```

Pada fungsi main, terdapat beberapa perintah yang digunakan untuk melakukan iterasi pengambilan ulasan pengguna, menampilkan progress scraping, hingga menyimpan output scraping ke dalam file berekstensi .xlsx.

```
def main():
    with sync_playwright() as p:
        browser = p.chromium.launch(headless=False)
        page = browser.new_page()
```

```

page.goto("https://maps.app.goo.gl/SyXMj8rCRsco4JM38", timeout=60000)

page.wait_for_timeout(5000)

page.locator('button:has-text("Ulasan lainnya")').click();

review_list = ReviewList()

page.wait_for_timeout(5000)

for i in range(1, total):
    review = Review()
    review_element =
page.query_selector('//*[@id="QA0Szd"]/div/div/div[1]/div[2]/div/div[1]/div/div/div[2]/div[9]/div['+str(3*i+-2)+']/div/div/div[2]/div[2]/div[1]/button')
    review_id = review_element.get_attribute('data-review-id')
    reviewer_name_xpath =
'//*[@id="QA0Szd"]/div/div/div[1]/div[2]/div/div[1]/div/div/div[2]/div[9]/div['+str(3*i+-2)+']/div/div/div[2]/div[2]/div[1]/button/div[1]'
    reviewer_name = page.locator(reviewer_name_xpath).inner_text()

    review_xpath = '//*[@id="'+review_id+'"]'

    if "... Lainnya" in page.locator(review_xpath).inner_text():
        button_lainnya_xpath =
'//*[@id="'+review_id+'"]/span[2]/button'
        page.locator(button_lainnya_xpath).click();

    review_text = page.locator(review_xpath).inner_text()
    review.id_review = review_id
    review.name = reviewer_name
    review.review_text = review_text

    review_list.review_list.append(review)

page.mouse.wheel(0, 7000)
page.wait_for_timeout(3000)
# Print 14 huruf terakhir review id dan total review yang sudah di
scrape
print(f"Review ID: ...{review_id[-14:]} | Currently Scraped: {i}",
end='\r')
sys.stdout.flush()

review_list.save_to_excel("D:/RAIHAN STIS/Perkuliahan/SEMESTER
5/Praktikum INFORMATION RETRIEVAL/Pertemuan
(12)/jakarta_aquarium_review_data")
browser.close()

```

Pada iterasi pengambilan ulasan pengguna, terdapat perintah 'page.mouse.wheel(0, 7000)' yang digunakan untuk mengatur gulir mouse secara otomatis. Nilai 0 menyatakan jumlah gulir secara horizontal dan nilai 7000 menyatakan jumlah gulir vertical sebanyak 7000 unit. Jumlah gulir vertical dapat disesuaikan dengan jumlah ulasan yang akan discraping.

```
page.mouse.wheel(0, 7000)
page.wait_for_timeout(3000)
# Print 14 huruf terakhir review id dan total review yang sudah di scrape
print(f"Review ID: ...{review_id[-14:]} | Currently Scraped: {i}", end='\r')
sys.stdout.flush()
```

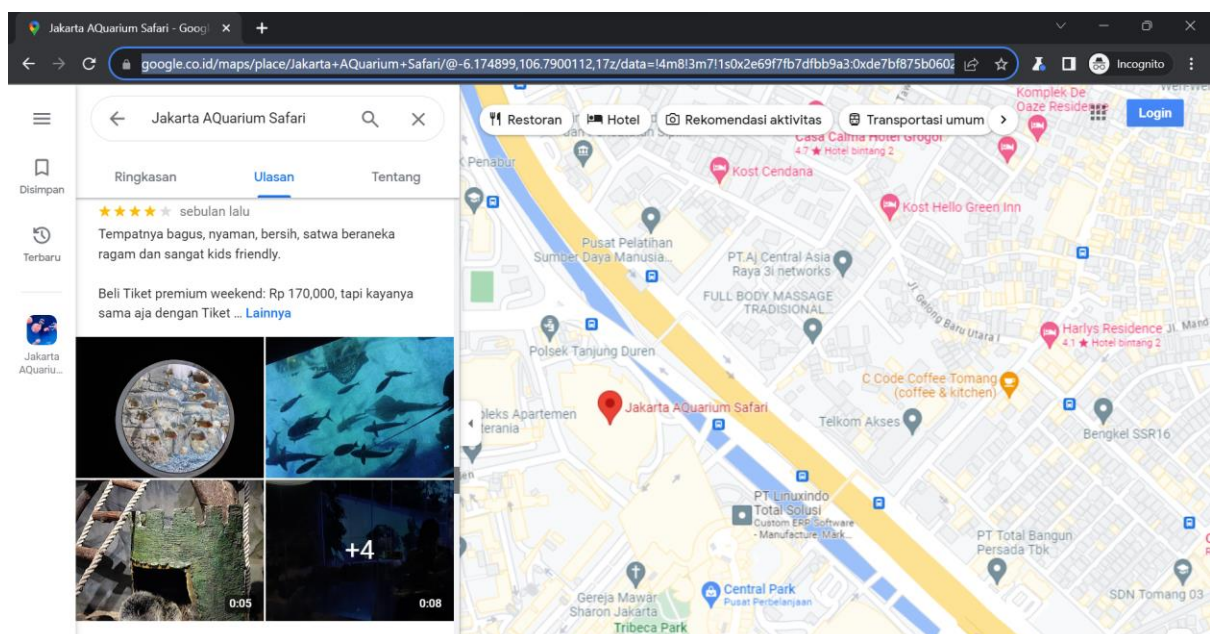
Program kali ini diatur untuk melakukan scraping total ulasan pengguna sebanyak 20 buah sehingga perlu dilakukan pendefinisian argument pada fungsi main.

```
if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument("-t", "--total", type=int)
    args = parser.parse_args()

    # total number of products to scrape. Default is 10
    if args.total:
        total = args.total + 1
    else:
        total = 20

    main()
```

Berikut merupakan tampilan output ketika kode program dijalankan.



```
(base) D:\RAIHAN STIS\Perkuliah\SEMESTER 5\Praktikum INFORMATION RETRIEVAL\Pertemuan (12)>python penugasan12.py -t 20
Review ID: ...NgcDVUdnNBRRAB | Currently Scraped: 4
```

```
(base) D:\RAIHAN STIS\Perkuliah\SEMESTER 5\Praktikum INFORMATION RETRIEVAL\Pertemuan (12)>python penugasan12.py -t 20
Review ID: ...UNacXZ10E9nEAE | Currently Scraped: 20
(base) D:\RAIHAN STIS\Perkuliah\SEMESTER 5\Praktikum INFORMATION RETRIEVAL\Pertemuan (12)>
```

	id_review	name	review_text
2	ChZDSUhNMG9nS0VjQ0FnSUNGdUpDZ	Nazifa	Beli tiket premium datang kesini pagi dan pengunjungnya cukup ramai karena ada kelompok study tour yang main ke sini.Khusus pengun
3	ChdDSUhNMG9nS0VjQ0FnSURGOXNQf	Bryce Orton	Tempat ini sangat seru. Ada banyak hewan untuk melihat dan dan juga bisa berinteraksi dengan beberapa. Anda bisa memegang Burung
4	ChZDSUhNMG9nS0VjQ0FnSURGNXITNI	Mirza Herlambang	Rupanya tempat yg populer bagi study tour anak TK/sekolah. Terletak di basement. Dari pintu masuk lobi yg ada Uniqlo, sebelah kiri langsu
5	ChdDSUhNMG9nS0VjQ0FnSUNGcDVUd	Tsanas Nabillah Sulistiyad	Experience untuk melihat binatang laut dan beberapa reptile, sayangnya untuk pertunjukan 5D sedang maintenance. But overall sangat s
6	ChZDSUhNMG9nS0VjQ0FnSUNaNmI3S	Nabiila Putri Caesari	baru pertama kali ke sini tempatnya cozy banget dingin dan lebih baik dari pagi supaya masih sepi jadi bisa puas foto foto dan kalau bisa i
7	ChdDSUhNMG9nS0VjQ0FnSURacGRyN	Benny Lin	Tempatnya bagus, tidak hanya hean lait, ada beberapa koleksi hewan darat juga dan Burung Macaw. Banyak Show, walau harga tiket aga
8	ChZDSUhNMG9nS0VjQ0FnSUNGnd6RI	Hwangza	Aku kesini pas weekend jd rame bgt, tp tetep seneng karena kesini berdua sama sayanku ave.Untuk tiket pas kesini 175k/person (weeker
9	ChdDSUhNMG9nS0VjQ0FnSUNGnJVybr	Sebrina Kartikasary	Kesini acara Fieldtrip sekolah anak, tempatnya bagus dan nyaman, jenis ikannya buanyakkk petugas guidenya jg ramah dan jelas bgt dlm m
10	ChdDSUhNMG9nS0VjQ0FnSURaeXJMSz	Tiara	aesthetic banget 🤩 cocok buat tempat liburan keluarga atau nge-date. Tempatnya ga besar karena dibawah mall tapi hewan hewannya t
11	ChdDSUhNMG9nS0VjQ0FnSURaX182Xz	Hawa Indah	Tempatnya bagus, nyaman, bersih, satwa beraneka ragam dan sangat kids friendly.Beli Tiket premium weekend: Rp 170,000, tapi kayanya
12	ChdDSUhNMG9nS0VjQ0FnSUNaef9hU	Muhammad Mufashih	Tempat hiburan dan edukasi yang sangat modern dan unik.Sangat cocok untuk wisata keluarga terutama yang punya anak-anak.Lebih dis
13	ChdDSUhNMG9nS0VjQ0FnSUNGd2FLZ	Maulita Pramesti	Dari awal adanya Jakarta Aquarium penasaran banget kaya apa.. ternyata rasa penasaran saya terjawab.Awalnya bingung aquarium dala
14	ChdDSUhNMG9nS0VjQ0FnSUM1NG9Ql	Ratih Pratitis	Akses mudah di tengah kota, dekat dengan halte busway. Menyenangkan sekali di sini, bisa lihat berbagai macam ikan dan hewan air lain
15	ChZDSUhNMG9nS0VjQ0FnSUNwdfTU	Ruli Retno	Tempatnya kecil aja sebenarnya, tp cukup mewakili keinginan tauhan anak2. Enaknya tuh lokasinya di mall dan buka sampai malam, jd bisa
16	ChdDSUhNMG9nS0VjQ0FnSURGMEktei	Ius Taat	Tempat yang bagus untuk wisata keluarga karena terdapat banyak poster yang mengedukasi, koleksi ikan dan binatang lainnya lumayan l
17	ChZDSUhNMG9nS0VjQ0FnSUNwd1oyV	- malfinhanif -	Tempat rekreasi edukasi untuk anak-anak yg berada di dalam kota jakarta. Bahkan berada di gedung yg bonafide. Anak-anak sangat sena
18	ChdDSUhNMG9nS0VjQ0FnSUMYaiRUU	Rere Rere	Tempat ya better sih drod ke ancol soalnya keiauhan. Tapi beli tiketnya ke website ya cek IG nya aia harga weekend dan weekdays beda so
19			