JOURNAL OF EVENT, TOURISM AND HOSPITALITY STUDIES
https://e-journal.uum.edu.my/index.php/jeth

How to cite this article:
Kamarazaman, N., Ali, N.M., & Arshad, H.. (2024). Leveraging web scraping to gather tourism information data. *Journal of Event, Tourism and Hospitality Studies*, 4, 16-29. https://doi.org/10.32890/jeth2024.4.2

LEVERAGING WEB SCRAPING TO GATHER TOURISM INFORMATION DATA

[1] Nadzirah Kamarazaman, [2] Nazlena Mohamad Ali & [3] Haslina Arshad

Universiti Kebangsaan Malaysia (UKM),
43600 Bangi, Malaysia

[1]Corresponding author: nadzirahk@gmail.com

Received: 9/11/2023 Revised: 24/3/2024 Accepted: 20/7/2024 Published: 31/7/2024

ABSTRACT

The influence of Information and Communication Technologies (ICT) on both individuals' daily lives and the economy is of significant importance. In this context, the tourism industry plays a crucial role, and it is essential to recognise the contributions of tourists in terms of sharing their experiences through tourism websites. Analysing this data is key to improving future tourists' experiences. Therefore, the objective of this study is to employ web scraping to gather data on places of interest (POI) and user attributes, specifically in the state of Melaka via the TripAdvisor website. Melaka is chosen as it is one of the places recognised by the United Nations, Educational, Scientific and Cultural Organization (UNESCO). The study focuses on the 200 POI locations (UNESCO) Map, encompassing both Melaka's core and buffer zones. These POIs are categorised into four heritage types: built heritage, natural heritage, personal heritage, and living heritage, with some belonging to more than one category. For the data collection process, this study utilised the TripAdvisor website and extracted a total of 14 attributes. Specifically, 27282 user data entries were collected from 163 POIs in the core zone area, and 8305 data entries from 37 POIs in the buffer zone area. The data is managed and stored in various formats, including CSV, JSON, and Excel files in the repository. The data helps in the development of a tourism application. Furthermore, the tourism industry can benefit from this study by enhancing their services and conserving the cultural heritage.

Keywords: Buffer zone, Core zone, Tourism data, UNESCO, Web scraping.

INTRODUCTION

The advancement in ICT has enabled tourists to effortlessly discover a myriad of travel destinations. With the emergence of new tourist spots and the expansion of the existing destinations, it is vital to ensure effective promotion of these destinations to remain attractive (Pratiwi, 2020). A rising number of global tourists are opting to independently seek information for their holiday destinations, diminishing their reliance on travel agencies. Information systems have provided the tourism industry with valuable data on tourism information and tourists' behaviors (Pratiwi, 2020). Websites such as Agoda, Hotel.com, and TripAdvisor provide the best recomme10.32890/jeth2024.4.1ndations about the places, rooms and price ranges. This information is useful and gives benefits, especially to tourists. TripAdvisor provides various services in the travel sector, with a primary focus on user-generated reviews and recommendations, which have been traditionally associated with the platform (Alaimo et al., 2020). Platform users engage by sharing their experiences, providing ratings for places, and commenting on various tourist attractions worldwide. These reviews have the potential to influence tourists' decisions when planning their trips. Founded in February 2000 as a travel search engine, TripAdvisor has evolved into one of the world's largest online travel sites, leaving a lasting impact on the tourism industry.

Trip Advisor also provides data about tourism places through comments and ratings from the users and also about the users itself such as trip type, country, date visited and many more. With over 1 billion reviews and opinions covering nearly 8 million businesses, TripAdvisor is a go-to platform for travellers seeking deals on accommodations, booking experiences, reserving tables at restaurants, and tracking down prominent nearby places (TripAdvisor Fact, 2022). TripAdvisor is one of the largest travel sites in the world that provides a platform for people to share itineraries or details about their journey. Tourists can also share their experiences during their stay in accommodations through the site. Therefore, the objective of this study is to utilise web scraping to gather data profiles on POI in Melaka from the TripAdvisor website. This method automatically extracts data from TripAdvisor which can be utilised by multiple parties without incurring any charges. Scraping and scraping processes are combined to extract the data from the website. The choice of the web scraping method allows researchers to efficiently access a large volume of data at once (Himawan et al., 2020).

The state of Melaka was chosen in this study because Melaka has been acknowledged by UNESCO as a World Heritage City in 2008 and tourism is one of the most important sectors that contribute to Melaka's economy (Zainal et al., 2021). The history of the colonial era is reflected in Melaka's unique culture, including its architecture, traditional clothing, and music, which showcase the city's rich heritage (Business & Research, 2023). Melaka is a state that promotes heritage tourism, and visiting historic and cultural sites is the most popular activity among tourists (Amir et al., 2015). This current study utilises the TripAdvisor website for various purposes, such as obtaining the comments about the places from tourists, the rating of the places that they visit and the kind of trip types the tourists experience. The analysis from this study indicates that ratings and comments from the tourists had an impact on the places they visited.

RELATED WORK

Tourism Information Data

Tourism consists of cultural, social, and economic aspects that involve the movement of a person for business or their own reason outside their places (Meng et al., 2022). The Internet has emerged as a crucial platform for disseminating tourism information and facilitating sales. The tourism industry is looking for the best strategies by developing a tourism website, introducing electronic commerce and providing

governance related to the tourism industries (Ma, 2020). Digital transformation plays a crucial role in the tourism sector, facilitating the distribution of products and services. Additionally, it allows both stakeholders and tourists to embrace the benefits of globalisation and societal changes (Ribeiro et al., 2022). Information technology in tourism has broadened along with time, not only based on computer reservation systems and global distribution systems or internet business per say, but a system of systems based on smart technologies (Koo et al., 2015). Information technology that follows the traveller's travelling pattern is the core of the travel context and the role of information technology is not only to revamp the competition in tourism industries but also to enlarge the tourism boundaries (Koo et al., 2015). The public is progressively embracing digitalisation, thanks to the widespread use of the internet. Moreover, ICT has gained significant prevalence due to its accessibility and effectiveness (Ribeiro, 2022). Therefore, technologies are essential for the development of a variety of economic sectors, and the tourism industry is no exception.

United Nations Educational, Scientific and Cultural Organization (UNESCO)

UNESCO contributes to peace and security by fostering international cooperation in the fields of education, culture, communication, science, and information (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021). This organization nurtures knowledge sharing to hasten mutual understanding and perfect knowledge of people's lives. UNESCO (2021) has established ten (10) criteria, and for a site to be included in the World Heritage List, it must meet at least one (1) of these criteria. The criteria undergo regular revisions to align with the evolving nature of the World Heritage concept (UNESCO, 2021). In late 2004, World Heritage sites were chosen based on six (6) cultural and four (4) natural criteria. The Operational Guidelines for the Implementation of the World Heritage Convention outline the singular set of ten (10) criteria. Below is the list of these ten (10) criteria:

1. To symbolise a pinnacle of human creative ingenuity.
2. To showcase a significant exchange of human values over time or within a cultural region, encompassing advancements in architecture, technology, monumental arts, town planning, or landscape design.
3. To serve as a distinctive or at least remarkable testament to a cultural tradition or a civilisation, whether existing or vanished.
4. To be an outstanding example of a type of building, architectural or technology ensemble or landscape which illustrates (a) significant stage (s) in human history.
5. To stand as an exceptional instance of a conventional human settlement, land-use, or sea-use that represents a culture or cultures, showcasing human interaction with the environment, particularly when it becomes vulnerable due to irreversible change.
6. To be directly or tangibly associated with events or living tradition, with ideas, or with beliefs, withs artistic and literary works of outstanding universal significance. (The committee considers that this criterion should preferably be used in conjunction with other criteria).
7. To contain superlative natural phenomena or areas of exceptional natural beauty and aesthetic importance.
8. To serve as notable examples representing crucial stages in the Earth's history, encompassing the record of life, significant ongoing geological processes in the formation of landforms, or noteworthy geomorphic and physiographic features.
9. To stand as exceptional examples representing crucial and ongoing ecological and biological processes in the evolution and development of terrestrial, freshwater, coastal, and marine ecosystems, as well as communities of plants and animals.

10. To comprise the paramount and noteworthy natural habitats crucial for the on-site preservation of biological diversity, including those housing threatened species of outstanding universal value in terms of science or conservation.

Web Scraping

Numerous data sources available on the internet serve various purposes, significantly influencing decision-making within companies and research endeavours. The extraction of data from websites on the internet is commonly referred to as web scraping or web crawling (Gunawan et al., 2019; Khder, 2021). Various web scraping methods have been developed in the past few years and the most traditional is the copy and paste method and manual analysis technique. This method is not practical when the user needs to scrap a large number of datasets and the user is bound to make mistakes (Khder, 2021). Other methods that are used to scrap is by using Regular Expression (Regex), Hypertext Markup Language Document Object Model (HTML DOM) and XPath. These methods are more complex and need some additional programmes before it can be executed (Gunawan et al., 2019). Web scraping process includes website analysis, website crawling and data organisation (Krotov & Tennyson, 2021). This process requires a person to have a piece of technical knowledge in order to run the web scraping process.There are many different ways to do the web scraping process, Ho, (2020) defines some of the web scraping tools as listed below:

- Browser plug-in tools: These plugins can be seamlessly integrated into web browsers, necessitating user involvement to designate the specific location of information they intend to scrape.
- Desktop applications: It is known for the friendly user interface and users can easily navigate the functions without HTML and cascading style sheets (CSS) knowledge.
- Programming languages: Programming languages such as Python, R and C# can be used in web scraping but it takes time to do.

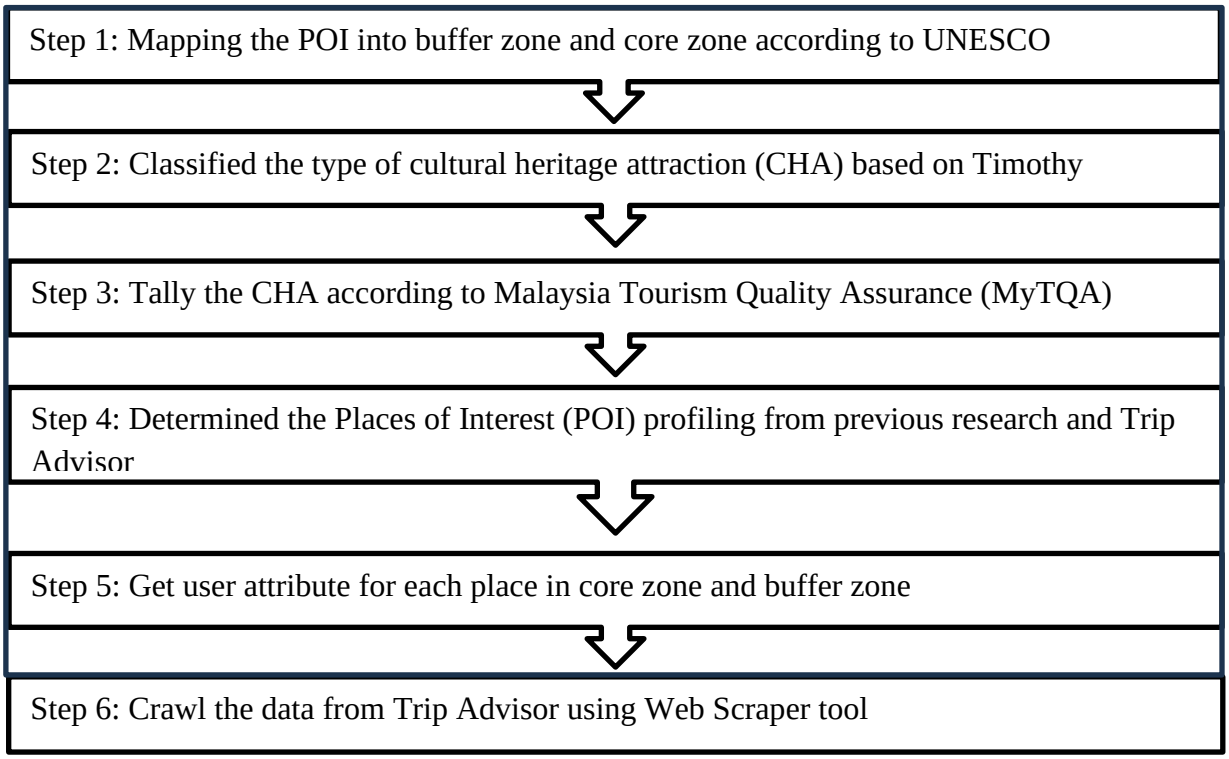
Nowadays, web scraping has become a main tool for statistical and scientific research of all types of research (Snell & Menaldo, 2016). The growing availability of open-source web scraping tools has significantly simplified the process for researchers, enabling them to construct customised web scrapers tailored to their specific needs (Han & Anderson, 2021). Web scraping and web indexing are interrelated, with web indexing utilising a web crawler to organise and catalogue information found on the web. Both positive and negative aspects are considered. The implementation comprises three elements which are a web crawler for fetching the links of the website, data mining to extract information from these links, and the storage of the acquired data in a CSV file (Sequeira et al., 2020). The internet offers numerous opportunities to comprehend consumer choices. Any information presented on travel websites serves as a potential data source for researchers in the hospitality industry (Han & Anderson, 2021). For instance, a researcher in the hospitality field could compile reviews from platforms like TripAdvisor and employ text-mining techniques. Alternatively, they may obtain satisfaction ratings for the purpose of conducting statistical analyses (Han & Anderson, 2021). Now, TripAdvisor holds the title of the world's largest travel website, featuring millions of traveller reviews and opinions. Widely acknowledged by consumers, TripAdvisor serves as a valuable resource for gathering insights before individuals plan their upcoming trips (Mahat & Hanafiah, 2020).

MATERIAL AND METHODS

The objective of this study is to implement the web scraping method using a Web Scraper application. The web scraping method automatically retrieves data from the TripAdvisor website without incurring any costs. This approach was selected because researchers are not required to use an Application Programming Interface (API), which often comes with access restrictions when retrieving data. Both qualitative and quantitative data were scraped in this study. This study utilised the web scraping method to extract data from the TripAdvisor website. The study involved six (6) distinct stages, outlined in Figure 1: Each step is discussed in detail in the following subsection.

Figure 1

The six stages of web scraping data for Melaka's core zone and buffer zone (Source: Author)

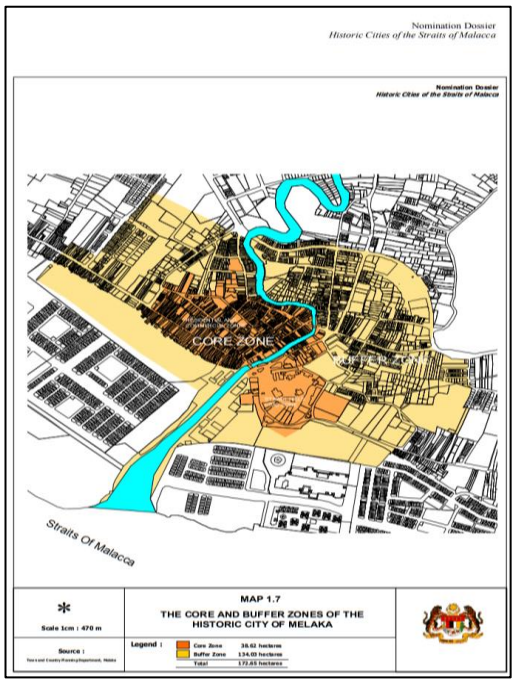


Step 1: Mapping the POI into buffer zone and core zone according to UNESCO

Melaka is divided into the core zone and buffer zone (Figure 2); the places of interest were manually added according to Melaka's map. The core zone area is located in the darker orange colour and the lighter one is the area for a buffer zone. Within Melaka City, the World Heritage Site consists of a core zone covering 38.62 hectares (15.7 percent) and a buffer zone spanning 134.03 hectares (84.3 percent) (Kelly Koh, 2017).

Figure 2

Core zones and buffer zones area in Melaka (Source: https://whc.unesco.org/en/documents/101085)



There is a total of two hundred places (200) from each core zone and buffer zone. One hundred and sixty-three (163) from the core zone and thirty-seven (37) places from the buffer zone. The places in both zones were manually tallied with the map to ensure accuracy.

Steps 2 and 3: Classified the type of cultural heritage attraction based on Timothy and tally the CHA to MyTQA

According to Timothy and Boyd (2003), there are different types and categories of cultural heritage. This study compared them with the Malaysia Tourism Quality Assurance (MyTQa) and presented the results in Table 1.

Table 1

Similarities of the type of cultural heritage and category of heritage between Timothy and Boyd (2003) with Malaysia Tourism Quality Assurance (MyTQa)

No.	Timothy and Boyd (2003)	Malaysia Tourism Quality Assurance (MyTQa)
1	Natural heritage: Natural Park and natural protected area	Natural Heritage: Nature and Adventure, Parks and Garden
2	Built heritage: Historic buildings, ancient ruins, and monuments	Built Heritage: Culture and Heritage
3	Living heritage: Food, customs and fashions	Living Heritage: Traditional Market / Bazars, Shopping, Tourism Restaurant, custom
4	Personal heritage: Cemeteries and religious sites	Personal Heritage: Place of Worship, cemeteries

Steps 4 and 5: Determined POI profiling from previous research and Trip Advisor and get user attributes for each place in the core zone and buffer zone.

This study scrapes both qualitative and quantitative data. On the TripAdvisor website, there are two (2) categories of data needed in this study which are first: POI profiling data including POI name, opening hours, address, contact number, description of the places, price, types of food (Asian, western, local, vegetarian, dessert, fusion) and types of meals (breakfast, brunch, lunch, tea-time, dinner, supper) as depicted in Figure 3, and second: user attribute data as illustrated in Figure 4 below:

Figure 3

POI Profiling Data on Trip Advisor Website (Source: Author)

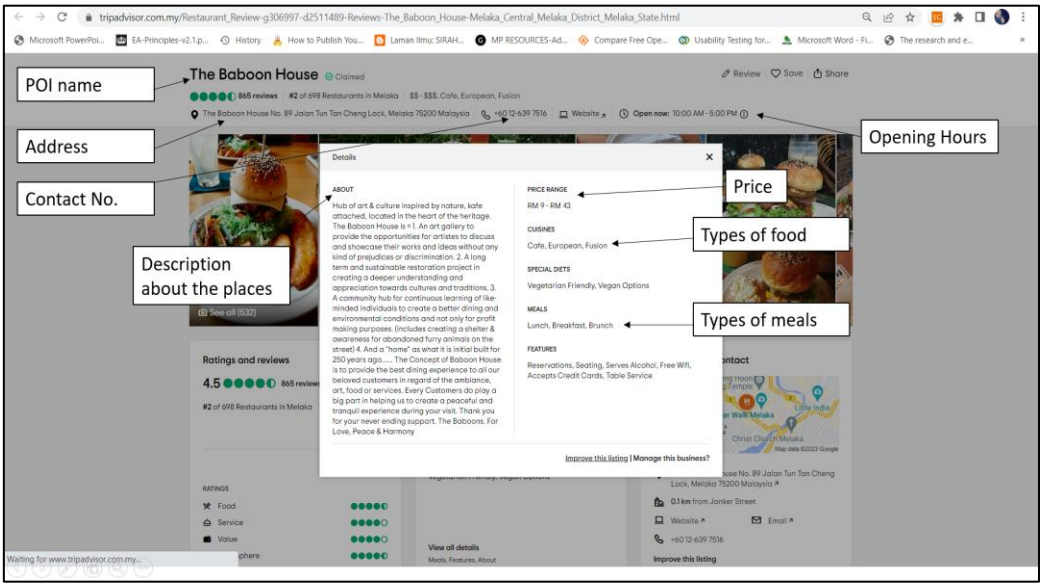


Figure 4

User attribute data for Core Zone and Buffer Zone places on the Trip Advisor website (Source: Author)

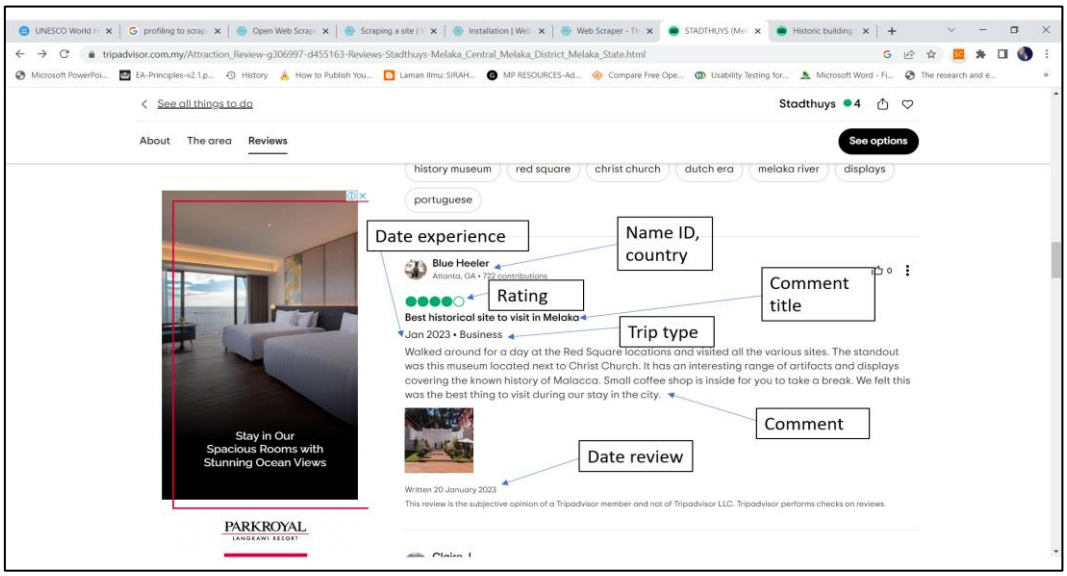
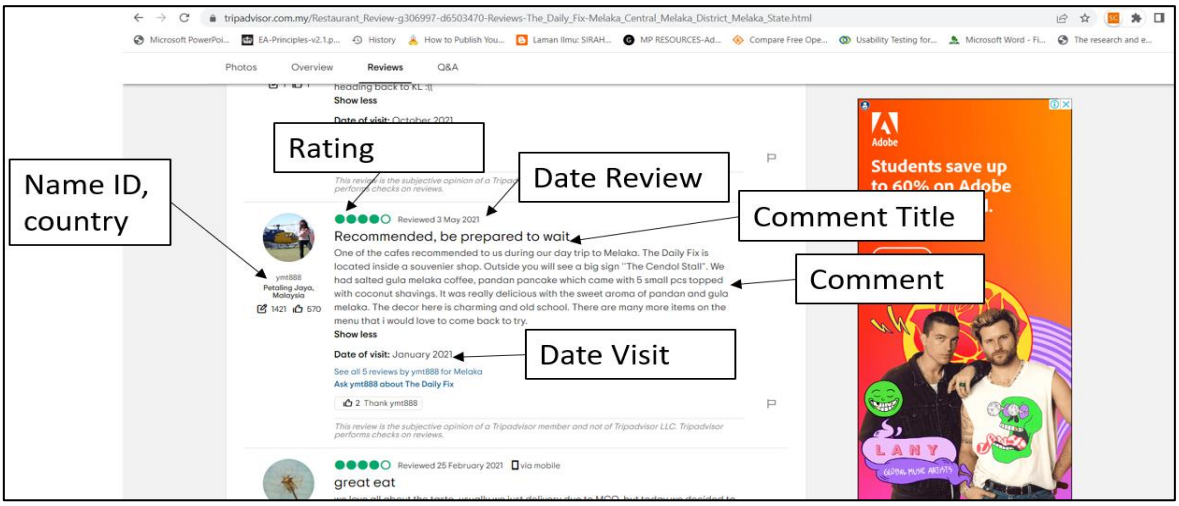


Figure 5

User attribute data for Core Zone and Buffer Zone restaurant on Trip Advisor website (Source: Author)



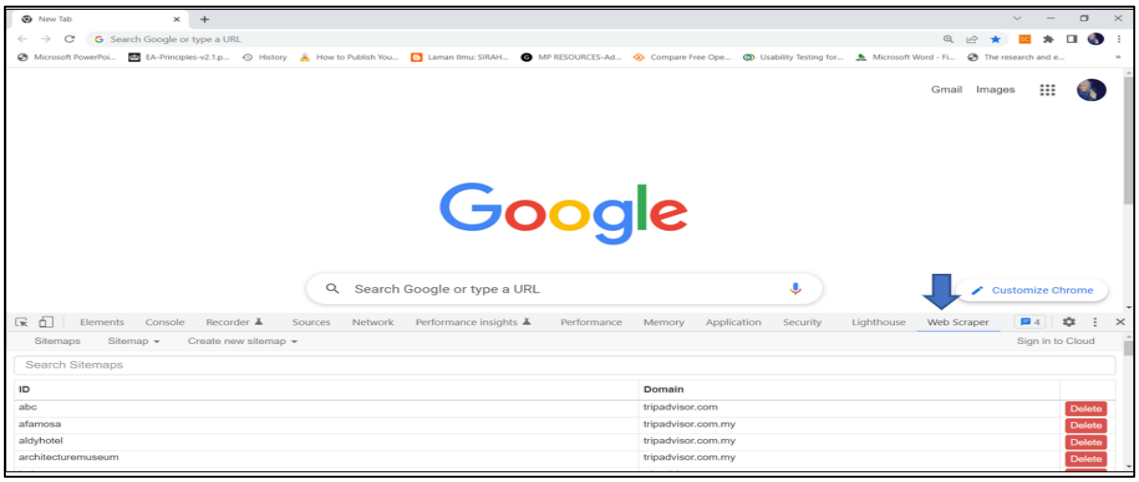
Step 6: Scraping the data from the TripAdvisor website using the Web Scraper tool

The online data extraction system, known as a web scraping tool, empowers users to access tourism data without requiring them to code manually. The data scraping technique involves the extraction of data from the internet (Himawan, Priadana, Murdiyanto, 2020). The extraction process is conducted to gather data from TripAdvisor. The steps involved in this study's extraction process, utilising the web scraping technique, are as follows:

- In the phase of POI profiling and obtaining user attributes, this study examines the HTML structure of the TripAdvisor website. The objective is to identify the elements that require scraping from the TripAdvisor website.
- The next process is to carrying out the crawl by using Web Scrapper software and this software can be integrated into browser developer tools as in Figure 6.

Figure 6

Web Scrapper integrated into browser development tool (Source: Author)



- data extraction process from TripAdvisor using Web Scraper is initiated by entering the HTTP request, which in this case is the TripAdvisor uniform resource locater (URL) (Figure 7). Next, select the specific element (POI Profiling and user attribute) on the Trip Advisor page that we intend to extract. Figure 8 illustrates the process of extracting data from the Trip Advisor website and the techniques used for scraping.

Figure 7

The URL page from Trip Advisor Website (Source: Author)

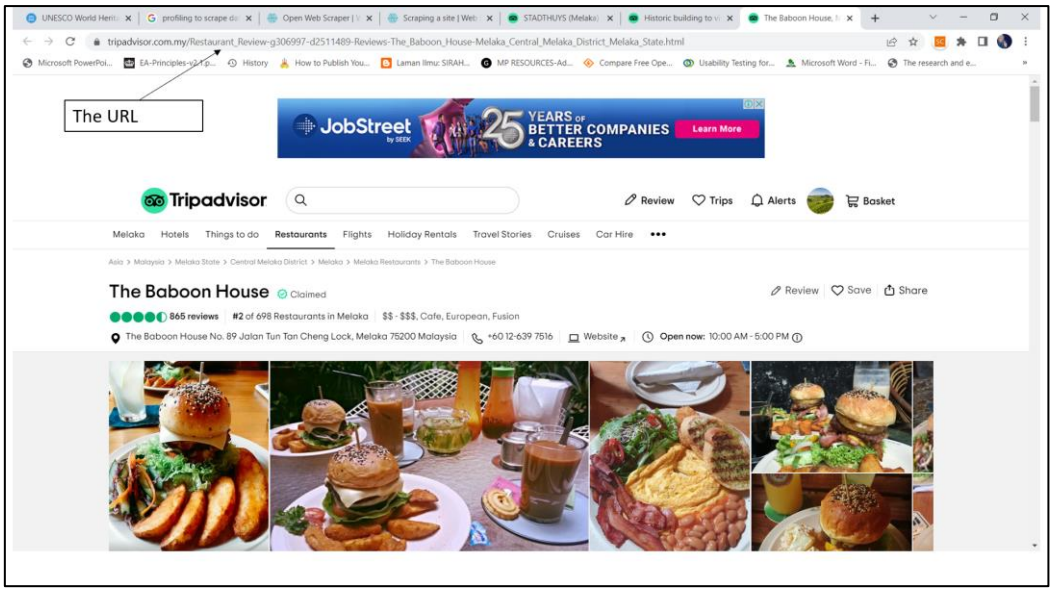
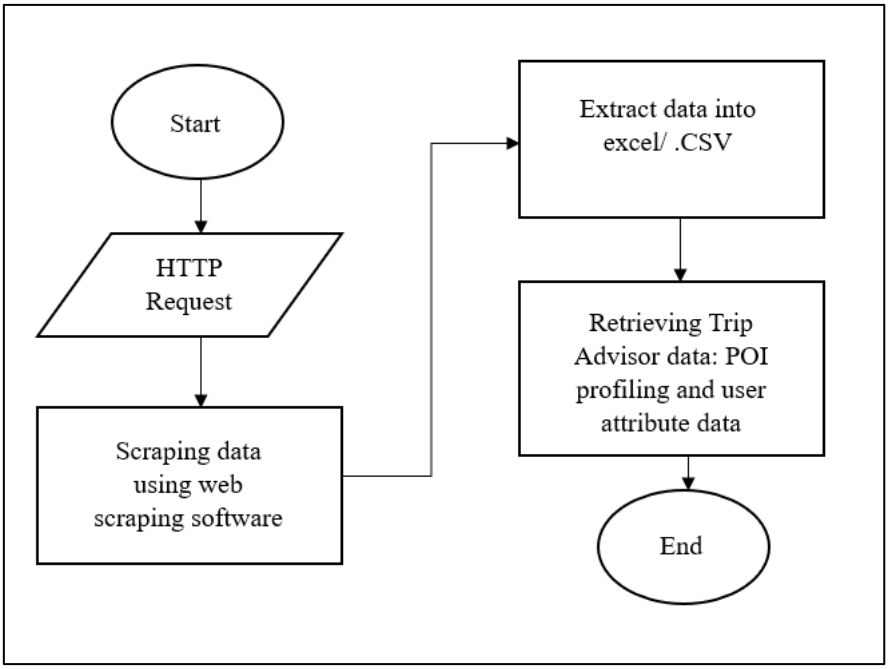


Figure 8

The process of extracting data from the Trip Advisor page (Source: Author)

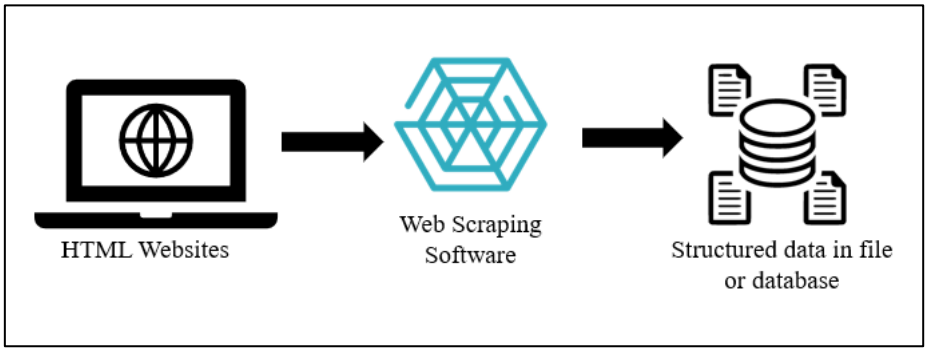


Application's Architecture Design

The architectural design of this application comprises three primary components: HTML websites, web scraping software, and structured data. Figure 9 illustrates the architectural design of this application.

Figure 9

The process of extracting data (Source: Author)



Each of these sections has detailed functions as follows:

- HTML Websites: This section functions as a media interface connecting users with web scraping software. Users have the capability to directly interact with various elements on a website. Additionally, through this module, users can oversee and manage the list of extracted data from TripAdvisor.
- A web scraper is a specialised tool specifically crafted to efficiently extract data from various websites. The design and complexity of web scrapers can vary significantly depending on the nature of the projects. Consequently, researchers should have a clear understanding of the specific data requirements associated with their work.
- Data Repository: This operates as a repository for data. In the architecture of this application, the approach employed for managing data in the repository includes formats such as CSV, JSON, and XML.

RESULTS AND DISCUSSION

This section highlights the results and discussion concerning the testing of the TripAdvisor data downloaded using the Web Scraper software as shown in Table 2. The upcoming evaluation aims to assess the efficacy of implementing the web scraping technique to download TripAdvisor data using black box testing methods (Himawan, Priadana, Murdiyanto, 2020). Prior to downloading the actual data, a preliminary trial was conducted to download the TripAdvisor data.

Table 2

Feature evaluation result

No	Features	Evaluation Result
1.	Downloading data without POI profiling and user attribute criteria	Success
2.	Downloading data of POI profiling criteria	Success
3.	Downloading with the number of user attribute criteria	Success
4.	Downloading with both POI profiling and user attribute criteria	Success

5.	Transfer downloading data to Excel	Success
6.	Transfer downloading data to CSV	Success
7.	Delete downloading data	Success

Number of POI according to the type of heritage

The total number of places that are involved in this scraping process is 200 POI which is located in the core zone (163 POI) and buffer zone (37 POI) area in Melaka (Table 3). From the scraping process, the POIs were categorised into 4 types of heritages according to Timothy and Boyd (2003) and Malaysia Tourism Quality Assurance (MyTQa). Some POIs belong to more than one (1) heritage category.

Table 3

Total number of POI according to the type of heritage

No	Type Of Heritage	Core Zone (Total) (C)	Buffer Zone (Total) (B)	Total (B+C)
1.	Built Heritage	23	11	34
2.	Natural Heritage	-	5	5
3.	Personal Heritage	4	2	6
4.	Living Heritage	131	18	149
5.	Built Heritage & Personal Heritage	5	1	6
	Total	**163**	**37**	**200**

From the table above, living heritage has the most POIs due to the abundance of hotels, shopping malls, and restaurants in Melaka. This is attributed to Melaka being a popular destination for both local and international tourists. There are numerous historical buildings from the Melaka Sultanate region in Melaka, which is why built heritage has the second highest number of POIs.

Scraping result

A total of 27282 user data from user attributes were collected from 163 POIs in the core zone area, while a total of 8305 data were collected from 37 POIs in the Melaka buffer zone area. The result is divided into four (4) types of heritage which are built heritage, natural heritage, personal heritage and living heritage. The total number of comments from users on the Trip Advisor website, across five satisfaction levels (excellent, very good, average, poor, and terrible), is represented by the result. The outcome is illustrated in Tables 4 and 5 below.

Table 4

The result of the web scrapping process from the Core Zone area

	Excellent	Very Good	Average	Poor	Terrible	Total Comment
Built Heritage	2514	2546	1106	137	62	6365
Natural Heritage	0	0	0	0	0	0
Personal Heritage	766	1105	507	46	12	2436
Living Heritage	6471	7097	3162	973	778	18481

					Total Comment	**27282**

Table 5

The result from the web scrapping process from the Buffer Zone area

	Excellent	Very Good	Average	Poor	Terrible	Total Comment
Built Heritage	286	472	333	59	24	1174
Natural Heritage	635	1143	555	112	88	2533
Personal Heritage	145	112	38	5	2	302
Living Heritage	2128	1414	485	149	120	4296
					Total Comment	**8305**

The final result is presented in the CSV file(s) containing the scrapped data entries where the results show that the POIs that fall under the categories of living heritage received the highest excellent satisfaction level in both the core zone and buffer zone areas. Hence, this study shows that tourists like Melaka because of the foods, bazaar and shopping area. The result of built heritage and personal heritage from the core zone area is mostly above average similar to the buffer zone area. The total comments from the core zone area are higher due to the fact that there are more POIs in the core zone area than in the buffer zone area. The natural heritage received a total of 1143 very positive comments from tourists. Tourism industries have the potential to enhance the satisfaction level of tourists by enhancing the natural heritage POIs. This result can benefit the businesses and tourism-related industries especially MyTQa to study the satisfaction level of the tourists about the POIs and heritage types. It will help business owners to cater to the users' satisfaction, especially in the living heritage category by maintaining their excellent services or upgrading their services.

CONCLUSION

This study has effectively employed the web scraping method to capture data from the TripAdvisor website. In this study, POI profiling data and user attribute data from the Melaka core zone and buffer zone areas were successfully downloaded with a total of 35587 user data from the user attribute. In this application, the data that have been downloaded can be managed and stored in CSV, Excel or JSON. The motivation behind commencing the web scraping process is to offer a resource for those who are engaged in the tourism industry. This study found that tourists were excellently satisfied with the living heritage POIs. Built and personal heritage also received an excellent and very good satisfaction level from the tourists. This result will help business owners or tourism industries boost their services, especially in terms of the living heritage POIs. Melaka's tourism industry can benefit from this study, where they can improve the built heritage by maintaining the building structure and making sure that the historical building is in a good condition. Future research can explore the potential of the tourism recommender application system for heritage sites, specifically targeting tourists. This will help tourists to explore and learn more about the heritage places. In the web scrapping process, some future research can be done to improve the web scraping process which is to enhance the scraping performance and efficiency by reducing the web scraping time because in this case, some places have a massive amount of data and it takes time to complete the scraping process.

ACKNOWLEDGMENT

The work was supported by university research KKP/2020/UKM/1 Konsortium Kecemerlangan Penyelidikan (KKP).

REFERENCES

Alaimo, C., Kallinikos, J., & Valderrama-Venegas, E. (2020). Platform evolution: A study of tripadvisor. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 5462–5471. https://doi.org/10.24251/hicss.2020.672

Amir, S., Osman, M. M., Bachok, S., & Ibrahim, M. (2015). Sustaining Local Community Economy Through Tourism: Melaka UNESCO World Heritage City. *Procedia Environmental Sciences*, *28*, 443–452. https://doi.org/10.1016/j.proenv.2015.07.054

Business, G., & Research, M. (2023). The Impact of Five Key Tourism Supply Chain Sectors on Tourists' Satisfaction of Cultural Heritage Tourism in Melaka, Malaysia. In *An International Journal* (Vol. 15, Issue 1).

Gunawan, R., Rahmatulloh, A., Darmawan, I., & Firdaus, F. (2019). *Comparison of Web Scraping Techniques : Regular Expression, HTML DOM and Xpath*. *2*(IcoIESE 2018), 283–287. https://doi.org/10.2991/icoiese-18.2019.50

Han, S., & Anderson, C. K. (2021). Web Scraping for Hospitality Research: Overview, Opportunities, and Implications. *Cornell Hospitality Quarterly*, *62*(1), 89–104. https://doi.org/10.1177/1938965520973587

Himawan, A., Priadana, A., & Murdiyanto, A. (2020). Implementation of Web Scraping to Build a Web-Based Instagram Account Data Downloader Application. *IJID (International Journal on Informatics for Development)*, *9*(2), 59–65. https://doi.org/10.14421/ijid.2020.09201

Ho, H. P. T. (2020). *Leveraging web scraping for collecting competitive market data: Case: A case study of an Airbnb rental unit in Helsinki*.

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications*, *13*(3), 144–168. https://doi.org/10.15849/ijasca.211128.11

Koo, C., Gretzel, U., Hunter, W. C., & Chung, N. (2015). Editorial : The Role of IT in Tourism. *Asia Pacific Journal of Information Systems*, *25*(1), 99–104. https://doi.org/10.14329/apjis.2015.25.1.099

Krotov, V., & Tennyson, M. (2021). *Web Scraping in the R Language : A Tutorial Journal of the Midwest Association for Information Systems Web Sc r aping in the R Language : A Tutorial*. *2021*(1). https://doi.org/10.17705/3jmwa.000066

Ma, H. (2020). The construction path and mode of public tourism information service system based on the perspective of smart city. *Complexity*, *2020*, 1–15. https://doi.org/10.1155/2020/8842061

Mahat, N. Z. D., & Hanafiah, M. H. (2020). Help me tripadvisor! examining the relationship between tripadvisor e-WOM attributes, trusts towards online reviews and travellers behavioural intentions. *Journal of Information and Organizational Sciences*, *44*(1), 83–112. https://doi.org/10.31341/jios.44.1.4

Meng, W. Y., Samad, A., Shibghatullah, B., & Subaramaniam, K. (2022). *Go . Travel – A Smart Tourism Guide Mobile Application*. *3*(2), 92–101.

Pratiwi, A. (2020). *Popular Destinations in Tourism: Case Study of TripAdvisor*. *426*(Icvhe 2018), 359–363. https://doi.org/10.2991/assehr.k.200331.167

Ribeiro, C., Silva, C. R. S., Azevedo, P. S., & Rosa, M. P. (n.d.). *View of Smart Travel Planning to the Algarve by Older Tourists before the Covid-19 Pandemic Crisis.pdf*.

Sequeira, S., Joy, J., Dsouza, D., & Kaul, P. (2020). Dynamic Review Modelling and Recommendation of Tourism Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3685669

Snell, J., & Menaldo, N. (2016). Web Scraping in an Era of Big Data 2.0. *Bloomberg BNA*, 2–5.

UNESCO. (2021). Operational Guidelines for the Implementation of the World Heritage Convention. *Operational Guidelines for the Implementation of the World Heritage Convention*, *WHS*, 188.

Zainal, Z., Merlimau, A. P., Latiff, H. A., Merlimau, P., Affyqah, N., & Merlimau, H. P. (2021). Tourists' Perceptions of Attraction Factors in Melaka 31 Politeknik &. In *Kolej Komuniti Journal of Social Sciences and Humanities* (Vol. 6, Issue 1).