ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

SCRAPING DATA FROM GOOGLE MAPS USING PYTHON

[1]TANINKI LAKSHMI SAI SARVANI, [2]P.BOBBY SOWJANYA

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

This project focuses on developing a web-based application to extract business information from Google Maps using Python-based web scraping techniques. The system enables users to search for business details such as restaurants, shops, or services by entering relevant queries. Upon receiving a query, the application connects to the internet and retrieves available data from Google Maps, including business names, ratings, reviews, addresses, and website links. The extracted information is then displayed in a structured format for easy understanding and analysis. The application is designed with user authentication modules, including user registration and login, ensuring secure access to the system. Once logged in, users can utilize the scraping module to fetch real-time business data based on their search queries. The system also allows users to explore additional details such as customer reviews and official websites of the businesses. The backend is implemented using Python, while MySQL is used for database management to store user information and query results efficiently. Although the system successfully retrieves useful business insights, it operates under certain limitations due to restrictions imposed by Google on free scraping services. As a result, only a limited number of results can be accessed without using paid APIs. Despite this limitation, the project demonstrates the practical application of web scraping and data extraction techniques in real-world scenarios. It can be further extended for applications such as market analysis, business intelligence, and location-based services.

Keywords: Web Scraping, Google Maps, Python, Data Extraction, Business Analytics, MySQL, Location-Based Services, Data Mining.

I.INTRODUCTION

In the modern digital era, location-based services and online business directories have become essential tools for users and organizations. Platforms like Google Maps provide vast amounts of information about businesses, including their location, ratings, reviews, and contact details. Extracting this

data can be highly beneficial for applications such as market analysis, competitor research, and customer behavior studies. This project focuses on utilizing web scraping techniques to collect business data from Google Maps using Python. By automating the data extraction process, the system reduces manual effort and enables users to access relevant information quickly and efficiently.

The core functionality of the system is based on web scraping, where Python scripts interact with web content to retrieve useful data. When a user enters a query, such as “restaurants near me” or any business-related keyword, the system connects to Google Maps and extracts relevant details. The retrieved data includes business names, ratings, reviews, addresses, and website links. This information is then displayed in a structured format for better readability. The system also integrates a database using MySQL to store user details and manage query results. This ensures efficient data handling and improves system performance. Despite restrictions on scraping by Google, the system demonstrates effective data retrieval within permitted limits.

The application is developed as a web-based platform with modules such as user registration, login, and data scraping. These modules ensure secure access and smooth functionality for users. The interface is designed to be simple and user-friendly, allowing users to easily perform searches and view results. Although the system currently retrieves limited results due to API restrictions, it can be extended using paid services or advanced scraping techniques. Overall, this project highlights the practical implementation of web scraping for extracting real-world data and its potential applications in business intelligence and analytics.

II SURVEY OF RESEARCH

The study by J. Leskovec, A. Rajaraman, and J. Ullman (2014) [1] explored large-scale data extraction and mining techniques from web sources. The methodology focuses on crawling and scraping structured and unstructured data from online platforms. Results showed that web data mining can significantly improve business intelligence and decision-making. However, challenges such as data inconsistency and legal restrictions were identified. This research provides a foundation for extracting business data from platforms like Google Maps.

The work by G. Salton and M. McGill (1983) [2] introduced information retrieval techniques for extracting relevant data from large datasets. The methodology involves indexing and querying mechanisms to retrieve accurate results. Results demonstrated improved efficiency in handling large-scale data queries. However, the approach requires optimization for real-time applications. This study supports

the query-based data extraction used in this project.

The study by M. Richardson (2007) [3] focused on web scraping techniques using automated tools and scripts. The methodology involves parsing HTML content and extracting specific data fields. Results showed that automated scraping significantly reduces manual effort and increases efficiency. However, website restrictions and dynamic content handling remain challenges. This research is directly related to implementing scraping functionality in the system.

The research by R. Kohavi (2001) [4] discussed the importance of data mining for business analytics. The methodology uses data collection, preprocessing, and pattern analysis to derive insights. Results indicated improved marketing strategies and customer targeting. However, data quality and privacy concerns were highlighted. This study supports the application of extracted data for business analysis.

The study by S. Brin and L. Page (1998) [5] introduced search engine mechanisms for ranking and retrieving web data. The methodology uses algorithms to rank web pages based on relevance and importance. Results showed highly efficient data retrieval systems. However, adapting these techniques for scraping requires careful implementation. This research is useful for understanding how data is organized in web platforms like Google Maps.

The work by E. M. Voorhees (2002) [6] explored evaluation techniques for information retrieval systems. The methodology includes precision and recall metrics to measure system performance. Results demonstrated effective evaluation of data extraction systems. However, maintaining high accuracy with dynamic data remains challenging. This study helps in evaluating the performance of the proposed scraping system.

III. WORKING METHODOLOGY

The proposed system follows a structured methodology for extracting business data from Google Maps using Python-based web scraping techniques. Initially, the system begins with user interaction through a web-based interface, where new users register and existing users log in securely. Once authenticated, users can access the scraping module by entering a query such as “restaurants,” “hospitals,” or any business-related keyword. The system then establishes an internet connection and sends requests to Google Maps to retrieve relevant search results. The raw data obtained from web pages is typically in HTML or dynamic content format, which is then processed using Python libraries. This stage ensures that the system collects accurate and relevant business information based on the user’s query.

In the next phase, the system processes and extracts useful information from the retrieved web data. Parsing techniques are applied to identify and extract specific fields such as business name, ratings, reviews, address, and website links. Libraries such as BeautifulSoup or Selenium are commonly used to handle static and dynamic content. The extracted data is then cleaned and structured into a readable format, removing unnecessary or duplicate information. This structured data is displayed to the user through the web interface, allowing easy interpretation. Additionally, the system enables users to view detailed information such as customer reviews and official websites of selected businesses. This step ensures that the extracted data is meaningful, organized, and useful for analysis or decision-making purposes.

Finally, the system integrates a database management component using MySQL to store user details and, optionally, query results. This ensures efficient data handling and improves system performance. The architecture is designed to handle multiple user requests while maintaining security and reliability. However, due to restrictions imposed by Google on free scraping, the system may return limited results unless advanced or paid APIs are used. Despite these limitations, the methodology demonstrates an effective approach for real-time data extraction from online platforms. The system can be further enhanced by incorporating advanced scraping techniques, automation, and cloud-based deployment for scalability and improved performance in real-world applications.

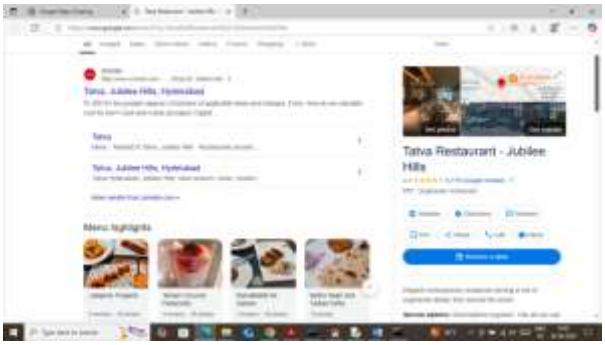
IV RESULTS EXPLANATIONS

In this project we are scrapping Google Maps to get all business details of any user queries and to implement this project we have designed following modules.

1) New User Register: user can sign up with the application
2) User Login: user can login to system
3) Scrape Google Maps: using this module user can enter any queries and then application will utilize internet connection to scan Google Maps to get all possible business details related to given query.

In above screen I am giving some query to get restaurants business details and then press button to get below page

In above screen got some search result for given query along with reviews, ratings, address and can click on 'View Website' link to view reviews and other details



In above screen for selected restaurant user can view all Google reviews and can view website details. Similarly enter any other query and get scrape output. in below screen trying another query

In above screen giving another query and below is the output

In above screen giving another query and below is the output

Similarly give query and get output

Note: as we know now-a-days Google services are not free of cost even for scrapping so when we search with free codes then Google will give few search results and in paid versions we can get many search outputs.

V.CONCLUSION

The proposed project successfully demonstrates the use of Python-based web scraping techniques to extract business data

from Google Maps. The system provides an efficient and automated way to collect useful information such as business names, ratings, reviews, addresses, and website links based on user queries. By integrating user authentication modules, the application ensures secure access and smooth interaction. The use of a web-based interface enhances usability, allowing users to easily perform searches and view structured results. This project highlights the importance of web scraping in real-world applications such as business intelligence, market research, and location-based analysis.

Despite its effectiveness, the system has certain limitations. The primary challenge is the restriction imposed by Google on free scraping, which limits the number of results that can be retrieved. Additionally, handling dynamic web content and ensuring data accuracy can be complex. The system also depends on internet connectivity and may face performance issues when dealing with large-scale data extraction. These limitations indicate the need for more advanced techniques and tools to improve scalability and reliability.

In future work, the system can be enhanced by integrating official APIs, cloud-based processing, and advanced data analysis techniques. Incorporating machine learning can further improve data filtering and recommendation capabilities. Overall, the project demonstrates a practical and scalable approach to extracting and utilizing web data for various applications.

RE.FERENCES

[1] J. Leskovec, A. Rajaraman, and J. Ullman, *Mining of Massive Datasets*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[2] G. Salton and M. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, 1983.

[3] M. Richardson, "BeautifulSoup documentation," 2007.

[4] R. Kohavi, "Data mining and visualization," in *Proc. ACM SIGKDD*, 2001.

[5] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw.*, vol. 30, pp. 107–117, 1998.

[6] E. M. Voorhees, "Evaluation of information retrieval systems," in *Proc. SIGIR*, 2002.

[7] A. B. Downey, *Think Python*. Sebastopol, CA, USA: O'Reilly Media, 2015.

[8] W. McKinney, *Python for Data Analysis*. Sebastopol, CA, USA: O'Reilly Media, 2017.

[9] M. Lutz, *Learning Python*, 5th ed. Sebastopol, CA, USA: O'Reilly Media, 2013.

[10] K. Aswani, "Web scraping using Python," *Int. J. Comput. Sci. Inf. Technol.*, 2019.

[11] J. F. DiMarzio, *Python Programming with MySQL*. New York, NY, USA: McGraw-Hill, 2002.

[12] D. Beazley and B. K. Jones, *Python Cookbook*. Sebastopol, CA, USA: O'Reilly Media, 2013.

[13] R. Mitchell, *Web Scraping with Python*. Sebastopol, CA, USA: O'Reilly Media, 2018.

[14] R. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, Univ. California, 2000.

[15] M. Grinberg, *Flask Web Development*. Sebastopol, CA, USA: O'Reilly Media, 2018.

[16] SeleniumHQ, "Selenium documentation," 2020.

[17] P. Barry, *Head First Python*. Sebastopol, CA, USA: O'Reilly Media, 2016.

[18] T. Oliphant, *Guide to NumPy*. USA: Trelgol Publishing, 2006.

[19] J. VanderPlas, *Python Data Science Handbook*. Sebastopol, CA, USA: O'Reilly Media, 2016.

[20] Google Developers, "Google Maps Platform documentation," 2023.

[21] A. Y. Ng, "Machine learning and data extraction," Stanford Univ., 2016.

[22] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[23] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[24] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.