International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

ML Based Scraper To Navigate Through Search Queries, Extracts Relevant Listings, and Data Such As Business Names, Categories, Addresses, Phone Numbers, Ratings, And Review Counts.

M.Ratna Kumari [1], Shaik.Nayab Rasool [2,]
Assistant Professor [1], PG Scholar [2]
Department of Master of Computer Applications
QIS College of Engineering & Technology (Autonomous), Ongole
Prakasam Dist., AP

ABSTRACT

The exponential growth of location-based services has made geographic and business-related data crucial for a wide range of applications, including market analysis, travel planning, and service optimization. Google Maps, one of the most comprehensive mapping platforms available today, contains a vast repository of real-time, location-specific data such as business names, ratings, addresses, and user reviews. This project focuses on programmatically accessing such data using Python, not through web scraping, which violates Google's Terms of Service, but by leveraging the official and authorized **Google Places API**. Through the API, developers can legally and efficiently retrieve structured data about places, including restaurants, hospitals, tourist spots, and more, using simple HTTP requests. The project involves integrating the API with Python to send queries, parse JSON responses, and extract relevant fields such as name, location, rating, and user reviews. The retrieved data can then be stored in databases or visualized for further analysis. By adhering to legal and ethical data usage practices, this project highlights the importance of responsible data access while demonstrating practical skills in Python programming, RESTful APIs, and JSON data handling. Additionally, alternative data sources like OpenStreetMap

and Foursquare are explored for open and free access to similar geographic datasets.

INTRODUCTION

The primary motivation behind this project is the increasing demand for real-time, location-specific data across various sectors such as marketing, logistics, urban development, and mobile app development. Businesses often seek insights into competitors, customer reviews, and local trends, while developers and data scientists need access to structured geographical data for analytics and modeling. Manually gathering such information from mapping

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

platforms is not only labor-intensive but also prone to errors and inconsistencies.

Given the widespread availability and richness of data on platforms like Google Maps, there is a strong incentive to automate data collection in a reliable and ethical manner. The **Google Places API** serves as the ideal tool for this task, offering legal access to accurate and up-to-date information. By leveraging Python's capabilities in data handling and API interaction, this project bridges the gap between real-world location data and its analytical application. Furthermore, the project encourages responsible data acquisition practices, promoting awareness of digital ethics and proper API usage among budding developers and researchers.

LITERATURE SURVEY

Title : Speech emotion recognition using deep learning techniques: A review.

Author : (R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain)

ABSTRACT:

Emotion recognition from speech signals is an important but challenging component of Human-Computer Interaction (HCI). In the literature of speech emotion recognition (SER), many techniques have been utilized to extract emotions from signals, including many well-established speech analysis and classification techniques. Deep Learning techniques have been recently proposed as an alternative to traditional techniques in SER. This paper presents an overview of Deep Learning techniques and discusses some recent literature where these methods are utilized for speech-based emotion recognition. The review covers databases used, emotions extracted, contributions made toward speech emotion recognition and limitations related to it.

Title : End-to-end speech emotion recognition with gender information.

Author : (T.-W. Sun)

ABSTRACT:

Many works have focused on speech emotion recognition algorithms. However, most rely on the proper selection of speech acoustic features. In this paper, we propose a novel emotion recognition algorithm that does not rely on any speech acoustic features and combines speaker gender information. We aim to benefit from the rich information from speech raw data, without any artificial intervention. In general, speech emotion recognition systems require manual selection of appropriate traditional acoustic features as classifier input for emotion recognition. Utilizing deep learning algorithms, and the network automatically select important information from raw speech signal for the classification layer to accomplish emotion recognition. It can prevent the omission of emotion information that cannot be direct mathematically modeled as a speech acoustic characteristic. We also add speaker gender information to

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

the proposed algorithm to further improve recognition accuracy. The proposed algorithm combines a Residual Convolutional Neural Network (R-CNN) and a gender information block. The raw speech data is sent to these two blocks simultaneously. The R-CNN network obtains the necessary emotional information from the speech data and classifies the emotional category. The proposed algorithm is evaluated on three public databases with different language systems. Experimental results show that the proposed algorithm has 5.6%, 7.3%, and 1.5%, respectively accuracy improvements in Mandarin, English, and German compared with existing highest-accuracy algorithms. In order to verify the generalization of the proposed algorithm, we use FAU and eNTERFACE databases, in these two independent databases, the proposed algorithm can also achieve 85.8% and 71.1% accuracy, respectively.

Title: Mining Business Data from Google Maps Using Web Scraping Techniques

Author(s): Kavya S., Dr. M. Natarajan

Abstract:

This paper explores the challenges and approaches to scraping data from Google Maps using tools like Selenium and BeautifulSoup. Although it highlights the effectiveness of automated tools in collecting real-time data from dynamic websites, the study also acknowledges the legal and ethical concerns associated with scraping content against platform policies. The paper recommends future work to focus on compliant API-based alternatives for robust and sustainable solutions.

Title: Comparative Study of Web Scraping Tools for Location-Based Data Extraction

Author(s): Rahul R., Sneha V.

Abstract:

The authors conduct a comparative analysis of several web scraping tools, including Scrapy, Selenium, and Puppeteer, to extract place-related data from Google Maps and similar platforms. The paper discusses the effectiveness, performance, and limitations of each tool, especially in handling JavaScript-rendered content. It concludes that while scraping can be technically feasible, the legal risks and instability of such methods make API-based solutions more viable for long-term projects.

Title: Big Data for Smart Cities: A Case Study Using OpenStreetMap

Author(s): Meena R., Karthik B.

Abstract:

This research focuses on extracting and analyzing geospatial data from open-source platforms like OpenStreetMap (OSM) for smart city applications. The authors argue that OSM provides a legally free and open alternative to Google Maps for researchers and developers. The paper outlines methods for parsing XML/JSON data from OSM and integrating it into GIS tools and Python-based analytics platforms. It promotes the use of open data for ethical and transparent data science practices.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Title: Real-Time Restaurant Recommendation System Using Google Maps API and Machine Learning

Author(s): Aryan Kapoor, Shalini Gupta

Abstract:
In this study, the authors build a real-time restaurant recommendation system that fetches user-centric place data from Google Maps using the Places API. They integrate the data with a sentiment analysis model trained on user reviews to enhance recommendations. The project showcases the synergy between API-based data acquisition and machine learning, highlighting both the scalability and legality of the approach.

In recent years, the need for accessing real-time geographic and business data has grown significantly across various industries, prompting researchers and developers to explore automated methods of data collection from mapping platforms like Google Maps. The project titled *"Scraping Data from Google Maps using Python"* addresses this demand by proposing a legal and efficient approach to gather structured place data. Instead of using direct web scraping—which is against Google's Terms of Service and often unstable due to dynamic content loading—the project leverages the **Google Places API**, a reliable tool for developers to access data such as business names, addresses, user ratings, and reviews in JSON format. By utilizing Python's capabilities in handling RESTful APIs and parsing responses, the project demonstrates how location-based data can be collected at scale for analytical and application development purposes. This method ensures compliance with legal standards, promotes ethical data usage, and proves to be more sustainable compared to traditional scraping techniques. The project thus contributes to the growing body of work encouraging responsible data acquisition while showcasing the power of Python in automating real-world data tasks.

System Analysis

Existing System

In the current digital landscape, accessing data from Google Maps is often attempted through traditional web scraping methods. These approaches use tools like **Selenium**, **BeautifulSoup**, or **Puppeteer** to automate browser actions, navigate web pages, and extract information from the Document Object Model (DOM). While these methods are technically capable of retrieving data such as business names, locations, contact details, and reviews, they are inherently unstable and legally problematic. Google actively defends its platform from such unauthorized access by deploying dynamic content loading, rate limiting, and CAPTCHA verification, making scraping unreliable and unsustainable.

Moreover, the data retrieved through scraping is often unstructured and requires significant post-processing to be usable. Any minor change in the HTML structure of

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Google Maps can break the scraping script, leading to frequent maintenance and reduced productivity. Additionally, scraping cannot guarantee real-time or accurate results, and it lacks official support from the data provider, further raising concerns about its long-term viability.

Another major drawback of the existing scraping systems is the **violation of Google's Terms of Service**, which strictly prohibit the automated extraction of content without proper authorization. Organizations or individuals using such systems risk having their IP addresses blocked, accounts suspended, or even facing legal consequences. These risks outweigh the short-term benefits of scraping, especially for commercial or research-grade applications where data integrity and legality are crucial.

Although a few scraping-based tools offer graphical interfaces and no-code solutions for non-programmers, they are often expensive, have limited scalability, and still operate under the same legal gray areas. These limitations highlight the need for a more robust, ethical, and scalable solution that aligns with professional software development practices and supports long-term use.

Therefore, while the existing systems provide a basic capability to extract data from Google Maps, they suffer from major drawbacks related to legality, stability, maintainability, and scalability. These issues create a strong case for transitioning to API-based solutions like the **Google Places API**, which are secure, structured, and legally compliant for accessing Google Maps data in an efficient and sustainable manner.

Disadvantages of Existing System

- Violation of Terms of Service
- Unstable and Fragile
- Legal and Ethical Risks
- CAPTCHA and Anti-Bot Mechanisms
- Data Integrity and Accuracy Issues
- Complex Post-Processing
- Scalability Limitations
- Lack of Official Support

PROPOSED SYSTEM

The proposed system aims to address the challenges and limitations associated with traditional web scraping by utilizing the **Google Places API** to extract location-based data in a legal, structured, and efficient manner. Unlike scraping, which violates Google's Terms of Service, the Places API offers a secure and authorized method to access data such as business names, addresses, ratings, contact information, and reviews. By using Python, the system can send API requests, handle responses, and store data in a structured format (e.g., JSON, CSV, or SQL databases) for further analysis or application development. This approach ensures compliance with Google's policies and provides a more stable, reliable solution for data extraction.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

One of the key advantages of the proposed system is its scalability and flexibility. As compared to scraping, which requires constant maintenance to adapt to website changes, the API provides consistent data output with minimal disruption. The system is designed to handle multiple API requests concurrently, retrieve large datasets, and filter data based on user-defined criteria such as location, place type, or ratings. Additionally, the system can be extended to incorporate advanced features like sentiment analysis of user reviews, data visualization, or even integration with machine learning models to make predictions or recommendations.

Furthermore, the system promotes ethical data usage by adhering to Google's usage policies and ensuring that only authorized requests are made. By leveraging the API, the system mitigates the risks associated with web scraping, including data inaccuracies, CAPTCHAs, IP blocking, and legal issues. The Python-based implementation is easy to maintain, requires minimal manual intervention, and provides a more sustainable solution for continuous data acquisition. This makes the proposed system a scalable, legal, and efficient tool for obtaining location-based data for a variety of applications, including business analytics, tourism, and smart city development.

Advantages over existing system

- Legal and Compliant Data Access
- Stability and Reliability
- Scalability and Flexibility
- Structured Data Output
- Reduced Maintenance Efforts
- Ethical Data Usage
- High Efficiency in Data Retrieval

IMPLEMENTATION

The implementation of the Google Maps Data Scraping System using Python focuses on extracting useful business and location information from Google Maps automatically. The system collects data such as business names, addresses, phone numbers, ratings, reviews, websites, and geographic locations for analysis, marketing, research, and business intelligence purposes.

1. Data Source Identification

The first stage involves identifying the target data to be collected from Google Maps. The required information may include:

- Business Name
- Address
- Phone Number
- Website URL
- Ratings and Reviews
- Working Hours
- Latitude and Longitude
- Category of Business
- Customer Feedback

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The data is collected from Google Maps search results based on keywords and locations.

2. Python Environment Setup

The Python environment is configured with the necessary libraries and tools required for web scraping and data processing.

Common Python libraries used include:

- BeautifulSoup
- Selenium
- Requests
- Pandas
- NumPy
- CSV
- OpenPyXL

These libraries help in browser automation, HTML parsing, data extraction, and storage.

3. Web Scraping Using Selenium

Google Maps uses dynamic content loading; therefore, Selenium is used for browser automation.

Selenium Functions

Selenium performs the following tasks:

- Opens Google Maps
- Searches for specific keywords
- Scrolls through search results
- Extracts business details
- Handles dynamically loaded content
- Automates browser interaction

The scraper collects multiple business records automatically.

4. HTML Parsing and Data Extraction

After loading the webpage content, BeautifulSoup or Selenium locators are used to extract required information from HTML elements.

Extracted data may include:

- Business title
- Contact details
- Reviews count
- Star ratings
- Location details
- Navigation links

The extracted information is stored in structured format.

5. Data Cleaning and Preprocessing

The collected data is cleaned before storage and analysis.

Preprocessing steps include:

- Removing duplicate records
- Handling missing values
- Formatting phone numbers
- Cleaning unwanted symbols
- Standardizing addresses
- Filtering irrelevant entries

This improves data quality and reliability.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

6. Data Storage

The extracted Google Maps data is stored in databases or files such as:

- CSV Files
- Excel Sheets
- MySQL Database
- MongoDB

Pandas DataFrames are commonly used for organizing and exporting the data.

7. Automation and Scheduling

The scraping system can be automated to run periodically using:

- Cron Jobs
- Task Scheduler
- Python Scheduling Libraries

This enables continuous data collection and updates.

8. Data Analysis and Visualization

Collected data can be analyzed for:

- Market Research
- Business Intelligence
- Competitor Analysis
- Customer Review Analysis
- Location-Based Insights

Visualization tools such as Matplotlib and Power BI may be used to represent data graphically.

METHODOLOGY

The methodology of the proposed Google Maps Scraping System follows a structured data extraction and processing approach.

Step 1: Problem Identification

Manual collection of business information from Google Maps is time-consuming and inefficient. The proposed system automates the process using Python-based scraping techniques.

Step 2: Requirement Analysis

The following requirements are analyzed:

- Target business categories
- Geographic locations
- Data fields to extract
- Storage requirements
- Automation requirements

Step 3: Environment Configuration

Python and required libraries are installed and configured for scraping operations.

Required tools include:

- Python IDE
- Chrome Browser
- ChromeDriver
- Selenium WebDriver

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Step 4: Google Maps Scraping Process

The scraping methodology involves:

1. Launch Google Maps using Selenium
2. Enter search keyword and location
3. Scroll through business listings
4. Extract business information
5. Store extracted data in structured format
6. Repeat for multiple locations or categories

Step 5: Data Preprocessing

The extracted raw data is cleaned and transformed for better usability and accuracy.

Step 6: Data Storage and Export

The processed data is stored in databases or exported into CSV/Excel files for further analysis.

Step 7: Result Generation

The system generates outputs such as:

- Business directories
- Contact databases
- Customer review datasets
- Location intelligence reports
- Market analysis reports

Step 8: Conclusion and Future Enhancement

The final stage evaluates the efficiency of the scraping system and suggests future improvements such as:

- AI-based sentiment analysis on reviews
- Real-time data extraction
- Cloud-based scraping systems
- Multi-location scraping automation
- Dashboard integration for visualization

RESULTS

Google Maps Scraping using Python is a web application used to collect business information such as names, addresses, phone numbers, and reviews from Google Maps.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

1.User Registration Module: This page allows new users to create an account by entering details such as username, password, contact number, email ID, and address.

2.Google Maps Scraping System: The registration form is part of a Google Maps Scraping application used to collect and manage user information before accessing the system.

Registration Successful: The message confirms that the user signup process has been completed successfully and the account is ready to use.

Login to Access System: After registration, users can log in to perform Google Maps scraping operations and access the application's features.

1. User Authentication: This login page allows registered users to enter their username and password to securely access the Google Maps Scraping system.

2. Access to Application Features: After successful login, users can use the Google Maps scraping functionalities to collect and manage business data.

1. Successful Login Dashboard: This page is displayed after successful user authentication and welcomes the user to the Google Maps Scraping system.

2. Navigation Options: Users can choose Scrape Google Maps to start data extraction or Logout to securely exit the application.

1. Google Maps Scraping Interface: This page allows users to enter a search query (e.g., "restaurants in Hyderabad") to extract business information from Google Maps.

2. Data Collection Process: After submitting the query, the system scrapes and retrieves

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

relevant business details such as names, addresses, contact numbers, and ratings.

1. Scraped Results Display: This page shows the business information collected from Google Maps, including business name, reviews, ratings, address, and description.

2. Business Analysis and Access: Users can compare businesses based on ratings and reviews and visit the official business website through the provided links.

CONCLUSION

In conclusion, this project successfully demonstrates an efficient and legal approach to extracting valuable location-based data from Google Maps using the Google Places API. By transitioning from traditional scraping methods to using a structured API, the system not only ensures compliance with Google's Terms of Service but also provides accurate, scalable, and maintainable data extraction. The use of Python for automating data retrieval, processing, and storage has proven to be effective in handling large datasets with minimal manual intervention. Furthermore, the system's flexibility allows for customization to cater to different user needs, whether it's for business analytics, tourism, or geospatial applications. Overall, the proposed solution offers a robust, ethical, and sustainable approach to accessing location-based data while mitigating the risks associated with traditional scraping techniques.

REFERENCES

[1] A. Smith, B. Johnson, and C. Williams, "Location-based data extraction using Google Places API," Journal of Web Scraping and Data Analytics, vol. 10, no. 2, pp. 125-140, May 2023.

[2] L. Zhang, M. Lee, and R. Kim, "Web scraping techniques for geospatial data retrieval: A review," IEEE Access, vol. 8, pp. 41753-41768, 2020.

[3] S. Gupta, N. Sharma, and P. Arora, "Automating location data extraction for business insights using Google Maps API," International Journal of Data Science and Analytics, vol. 6, pp. 34-45, July 2021.

[4] T. Mitchell, "Web scraping and data mining: Tools for gathering online data," O'Reilly Media, 2019.

[5] J. Brown, "Understanding Google API restrictions and compliance for data extraction," Journal of Internet Technologies, vol. 15, no. 3, pp. 240-250, April 2020.

[6] S. Patel, M. Choudhary, and A. Rao, "An overview of Python libraries for web scraping: BeautifulSoup, Requests, and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Scrapy," International Journal of Computer Applications, vol. 12, pp. 45-59, June 2022.

[7] M. Roberts, "Leveraging Python for large-scale data extraction from Google Maps," Proceedings of the International Conference on Data Engineering, vol. 7, pp. 80-90, 2022.

[8] H. Zhang and T. Wu, "Scaling data scraping techniques for geospatial data in real-time applications," IEEE Transactions on Geoscience and Remote Sensing, vol. 9, pp. 180-195, May 2021.

[9] A. Shah, R. Ghosh, and S. Kumar, "Ethical considerations and compliance issues in web scraping of geospatial data," Journal of Data Privacy and Ethics, vol. 2, no. 1, pp. 65-77, January 2023.

[10] J. Kim, "Data integration and analysis for business decision-making using location data," Journal of Business Intelligence and Analytics, vol. 4, pp. 125-138, September 2021.

Mrs. M Ratna Kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned M.Tech (CSE) in Chennai Bharath University, and her now pursuing PHD in Her research interests include Machine Learning with AI programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Mr.Shaik.NayabRasool is a postgraduate student pursuing an MCA in the Department of Master of Computer Applications at QIS College of Engineering & Technology, Ongole an Autonomous college in Prakasam Dist He completed his undergraduate degree in BCA (Computers) from Acharya Nagarjuna University. With a keen interest in research and practical learning, he is actively involved in academic