Project Documentation for Semantic Analysis Application

Made by: Sushant Suman

College: Institute of Engineering and Management, Kolkata

Branch: Information Technology

Batch: 2020- 2024

Email: sushantsuman90@gmail.com

Submission for Telaverge Communications Recruitment Drive

Submission Date- 17/11/23

Introduction

Welcome to the documentation for the Social Media Analytics Application. This project is a powerful tool designed to collect and analyze data from various social media platforms, including Facebook, Instagram, and Twitter. By scraping valuable information such as likes, followers, reviews, and comments, the application provides insights into the online presence of a brand.

Purpose and Goal

The primary goal of this project is to empower users to monitor and assess the performance of their brands across different social media channels. The application utilizes web scraping techniques to extract relevant data, calculates a comprehensive score based on predefined criteria, and notifies users of any significant changes in the brand's online presence.

Key Features

- Multi-Platform Data Collection: The application supports data collection from major social media platforms, including Facebook, Instagram, and Twitter.
- Scoring System: A sophisticated scoring algorithm evaluates the collected data to provide a comprehensive performance score for the brand.
- Change Notification: Users receive timely email notifications when there are notable changes in the brand's online metrics, accompanied by an explanation of the factors contributing to the change.
- Configurability: The application is highly configurable, allowing users to customize API credentials, scoring parameters, and email notification settings.

How It Works

- Data Collection: The application accesses social media platforms using python libraries like BeautifulSoup and Selenium to retrieve information such as likes, followers, reviews, and comments.
- Scoring Algorithm: A scoring system processes the collected data, and assigns a general score of 10. When there is a change in online performance of the brand the change is reflected in score proportionately either it is positive or neagative.
- Change Detection: The system stores brand's social media metrics and identifies significant changes in the calculated score.
- Email Notification: Upon detecting a noteworthy change, the application sends an email notification to the configured address, detailing the change and the reasons behind it.

Usage

- Change the email written in **email_config.txt** to the email address on which notifications will be received.
- Use the **config.txt** file to set the social media sites from which the data will be extracted. (Presently only facebook and instagram can be used at maximum).
- Run the **main_program.exe** file to start the application.
- Enter the name of the brand whose data is to be collected.(Make sure to use the name that is used on social media websites)
- A new .txt file is created where all the data is stored and base score of 10 is set.
- All the collected data will be shown on the console and program stops.
- Run the program again and enter the same brand name, if there is any change in the data collected compared to previously stored data then the score will be altered and reasons for the change will be mailed to the configured Email ID.

Supported Social Media Platforms

Facebook

- From facebook the number of likes and followers from brand page is collected.

Instagram

- From Instagram the number of followers from the brand page is collected.

(More social media platforms to be added)

Working of the Application

Beautiful Soup is a Python library designed for web scraping purposes to pull the data out of HTML and XML files. It provides Pythonic idioms for iterating, searching, and modifying the parse tree, making it easy to navigate and manipulate the data within HTML or XML documents.

Key Features:

1. Parsing HTML and XML: Beautiful Soup transforms a complex HTML or XML document into a navigable tree structure, allowing easy traversal and extraction of information.
2. Search and Filter: The library provides powerful methods for searching and filtering the parse tree based on tags, attributes, or textual content, simplifying the extraction of relevant data.
3. Tag Navigation: Beautiful Soup supports tag-based navigation, enabling users to move up, down, sideways, and across the parse tree to access specific elements or their contents.
4. Modifying the Parse Tree: In addition to extraction, Beautiful Soup allows modifications to the parse tree, such as adding, removing, or modifying tags and their attributes.
5. Integration with Parsers: Beautiful Soup supports various parsers, including the built-in Python parser (html.parser), lxml, and html5lib, giving users flexibility in choosing the parsing method that best fits their needs.

Selenium is a robust open-source framework designed for automating web browsers, making it a powerful tool for web application testing and browser automation. It provides a suite of tools and libraries that allow developers to interact with web browsers programmatically, mimicking user actions and validating the behavior of web applications.

Key Features:

1. Cross-Browser Compatibility: Selenium supports multiple web browsers, including Chrome, Firefox, Safari, and Edge, enabling developers to create browser automation scripts that work seamlessly across different platforms.

2. WebDriver API: The WebDriver API in Selenium allows the automation of browser actions such as clicking, typing, navigating, and interacting with web elements. It provides a standardized interface for controlling browsers programmatically.

3. Testing Framework Integration: Selenium integrates well with popular testing frameworks such as JUnit and TestNG, making it an essential tool for implementing robust and scalable test suites.

4. Dynamic Page Interaction: Selenium handles dynamic web pages efficiently, allowing developers to wait for specific elements to load before executing actions. This ensures reliable automation in scenarios with asynchronous content loading.

5. Headless Browser Support: Selenium supports headless browsers, allowing users to run browser automation scripts without a graphical user interface. This is beneficial for running tests in environments without a visible desktop.

- Selenium has been used to reach the social media websites and extract the page source file.
- BeautifulSoup library has been used to find the required tags based on class names from the html source and extract the data needed.
- The data collected is then stored in brand.txt file and a base score is given.

Scoring

- All the data collected is in numeric form so they are all summed up to a single value and that value is considered the base value.
- When there is any change any of the values the change is reflected in the summed up value, the difference from old value is used to find percentage change in the sum.
- The percentage change calculated is used to make similar percentage change in the base score (10).
- For example if the change in the total is sum of -10% then the new score will be 9.

Future Improvements:

The future plan for the project is to provide more social media websites access to the system and extract more and precise data from the websites.

Conclusion

The Social Media Analytics Application is a comprehensive tool designed for users to monitor and evaluate their brand's performance across various social media platforms, including Facebook and Instagram. Leveraging web scraping techniques with libraries like BeautifulSoup and Selenium, the application collects crucial data such as likes, followers, reviews, and comments. It then employs a sophisticated scoring algorithm to calculate an overall performance score for the brand.