



Team 5
Nikita Makurin
Kirill Tolkunov
Yehor Tereshchenko
Miikka Lahtinen
Darijus Seporaitis

AI-Powered Research and Client Outreach Platform

Metropolia University of Applied Sciences, Aalto University

Code documentation

25.11.2024

Contents

1	Introduction	3
2	Description of the AI-Powered Research and Client Outreach Solution	3
2.1	Key Features	3
2.2	Workflow	4
2.3	Use Cases	5
3	Conclusion	5

1 Introduction

The MVP in form of web application is an implementation of the first step of the development of SmartLab's environment for relevant to construction field companies and independent researchers. This advanced platform that combines web scraping, data processing, and artificial intelligence to deliver insightful analyse. It is built using Python and leverages the Flask framework for its backend, with integration of the Gemini generative AI model for natural language processing tasks. This application is designed to streamline the process of gathering, analyzing, and presenting data for opportunities seeking actionable insights.

2 Description of the AI-Powered Research and Client Outreach Solution

In the following sections will be provided main details about the solutions, i.e. key features, workflow details, key technologies, use cases and benefits.

2.1 Key Features

This web application combines Flask-powered backend processing, data analysis, and an user interface to deliver an MVP of a comprehensive solution for SmartLab. It functions as a web-based platform, hosting essential static files such as index.html and JavaScript assets. These static files form the foundation of the UI that bridges backend operations with frontend visualization, ensuring smooth and accessible user interactions.

The application's data processing capabilities are centered around the /process endpoint, which handles user-submitted JSON data containing company details. This endpoint initiates a multi-step workflow: websites are scraped for relevant information, and the extracted content is analyzed using an AI model to generate actionable insights and targeted data. By streaming results incrementally, the platform enhances the user experience with real-time feedback, making large-scale operations more interactive and efficient.

Its website scraping functionality is powered by the BeautifulSoup library, allowing the application to extract required content and raw HTML from websites. Users can input a base URL, prompting the application to recursively navigate linked pages within the same domain. This ensures a thorough yet focused data collection process, with mechanisms to limit the number of pages scraped, optimizing efficiency and resource utilization.

A standout feature of the application is its integration with the **Gemini AI model**, which elevates the scraped data into insightful analyses and sales leads. Using pre-defined and editable prompts stored in a YAML file, the AI analyzes the content and formats its findings into structured HTML reports. This integration transforms raw data into actionable outputs, providing detailed insights that are directly derived from their target websites. Together, these features position the application as a powerful tool for SmartLab to streamline their data collection and analysis processes.

2.2 Workflow

The process begins with user interaction on the frontend and ends with delivering analyzed data. Users can upload a CSV file containing company details or submit JSON data through the /process endpoint. The application interface, hosted as a static file (index.html), serves as the primary point of access for these operations.

Once data is submitted, the backend performs web scraping. It navigates the websites listed in the input, extracting structured content while following same-domain links to ensure relevant data collection. This content is stored for further analysis.

The scraped data is processed by the Gemini AI model, which uses prompts configured in YAML files. These prompts adjust to the input, enabling the AI to produce meaningful analyses. The outputs are structured for clarity and usability.

The application streams results incrementally, allowing users to view updates in real-time. Alternatively, the output can be downloaded as a JSON file. Errors during processing are handled efficiently, skipping problematic entries without disrupting the overall workflow.

2.3 Use Cases

The platform supports SmartLab's needs by automating key processes such as generation of insights and market research. By analysing companies' websites, the application identifies potential clients and generates development ideas tailored to specific industries. This functionality enables SmartLab to expand its outreach and target relevant audiences effectively.

In addition to development enhancement, the platform provides tools for market research. It extracts insights about companies, including their industry focus and key offerings. This data aids Metropolia in understanding market trends and positioning themselves strategically.

The platform also facilitates content analysis, streamlining the examination of website materials for SEO and digital marketing purposes. By integrating AI-driven insights, Metropolia, ABB and Skanska can refine their strategies and make data-informed decisions. These capabilities, combined with enterprise intelligence features, allow organizations to leverage web data for broader strategic planning.

3 Conclusion

This web application is an innovative tool for the SmartLab aiming to leverage technology for competitive advantage. By integrating web scraping, AI-powered analysis, and real-time feedback mechanisms, it simplifies complex workflows and delivers actionable insights efficiently.