

Assessment Report: Web Research Agent

Poornima MC - <https://github.com/POORNIMA-MC/web-research-agent>

1. Objective

The objective of this project is to develop a Web Research Agent that can automatically:

- Accept a user query,
- Search the web for relevant information using SerpAPI,
- Scrape web content from the top search results,
- Clean and summarize the extracted content using transformer-based models,
- And display the summarized information neatly on a web interface built using Flask.

The goal is to save users time by providing condensed, accurate, and relevant summaries from multiple sources without the need to manually browse through various articles.

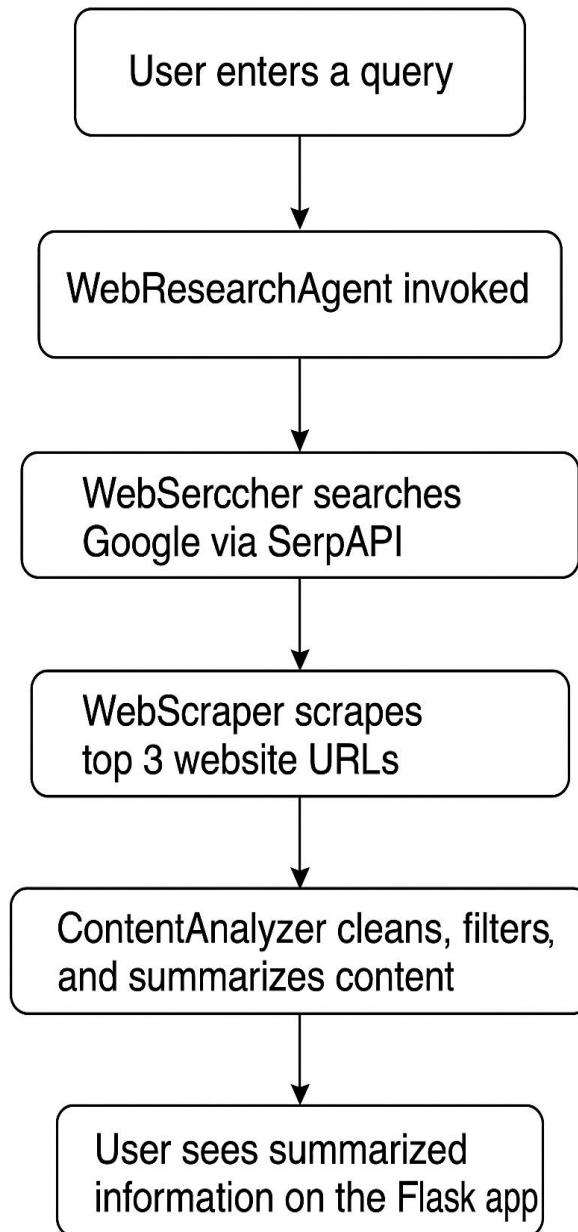
2. System Architecture and Flowchart

2.1 Architecture Overview

The system is composed of the following major components:

- **Flask Web Interface** for user interaction,
- **WebSearcher** to perform search operations using SerpAPI,
- **WebScraper** to extract website content,
- **ContentAnalyzer** to clean and summarize text.
- **WebResearchAgent** as the orchestrator of all components.

2.2 Flowchart



3. Module-wise Explanation

3.1 Flask Web Application

- Accepts user queries through an HTML form.
- Handles both GET and POST methods.
- Displays both the user's query and the resultant summarized content.

3.2 WebSearcher (Using SerpAPI)

- Takes a user query as input.
- Performs a Google search using SerpAPI.
- Extracts and returns top 3 organic result URLs.

3.3 WebScraper (Using Selenium and BeautifulSoup)

- Loads each URL in a headless Chrome browser.
- Extracts all `<p>` (paragraph) tags from the HTML.
- Joins all paragraph texts into a single raw text body.

3.4 ContentAnalyzer (Cleaning and Summarization)

- Cleans the scraped content by removing unwanted boilerplate texts (e.g., "contact us", "terms of service").
- Filters relevant sentences based on keyword matching with the query.
- Uses Hugging Face's BART-large-cnn model to summarize large extracted content.
- If the extracted text is already short and relevant, it is returned directly without summarization.

3.5 WebResearchAgent (Main Controller)

- Acts as the central coordinator.
- Manages the flow:
 - Search → Scrape → Analyze → Return summarized results.

4. Results

You: Jawaharlal Nehru

https://en.wikipedia.org/wiki/Jawaharlal_Nehru
Jawaharlal Nehru was an Indian anti-colonial nationalist, secular humanist, social democrat, and statesman. He was a principal leader of the Indian nationalist movement in the 1930s and 1940s. Upon India's independence in 1947, he served as the country's first prime minister for 16 years. Nehru promoted parliamentary democracy, secularism, and science and technology during the 1950s, powerfully influencing India's arc as a modern nation. He died in 1964 at the age of 83. He is survived by his wife, three children, and six grandchildren. His son Motilal was a prominent lawyer and Indian nationalist. His grandson is former Prime Minister of India Ravi Agrawal, who died in 2011.

<https://www.britannica.com/biography/Jawaharlal-Nehru>
Jawaharlal Nehru was the first prime minister of India after it gained independence. He established parliamentary government and became noted for his neutralist (nonaligned) policies in foreign affairs. He was born to a family of Kashmiri Brahmins, noted for their administrative aptitude and scholarship, who had migrated to Delhi early in the 18th century. On leaving Cambridge he qualified as a barrister after two years at the Inner Temple, London, where in his own words he passed his examinations "with neither glory nor ignominy." The seven years Nehru spent in England left him in a hazy half-world, at home neither in England nor in India. He died on May 27, 1964, in New Delhi.

https://www.pmindia.gov.in/en/former_pm/shri-jawaharlal-nehru/

Type your message...

Send

You: Who won Cricket ICC Champions Trophy 2025

https://en.wikipedia.org/wiki/2025_ICC_Champions_Trophy
The 2025 ICC Champions Trophy was the ninth edition of the tournament. It was hosted by Pakistan Cricket Board from 19 February to 9 March 2025. Co-hosts Pakistan were the defending champions and were eliminated in the group stage. India became the champions by defeating New Zealand in the final. They also became the first team to win three Champions Trophy titles. The tournament featured 15 matches held across three venues in Pakistan and one in the United Arab Emirates. The top two teams from each group advanced to the knockout stage, featuring two semi-finals leading up to the final and reserve days. The India–Pakistan cricket rivalry has been severely impacted by the tense political relations between the two nations. The PCB's plan to take the trophy to cities in Pakistan-administered Kashmir was objected to by the BCCI.

<https://www.icc-cricket.com/tournaments/champions-trophy-2025>

<https://www.espnricinfo.com/series/icc-champions-trophy-2024-25-1459031>

Type your message...

Send

5. Conclusion

The Web Research Agent effectively automates web research by integrating:

- Google search automation,
- Website scraping,
- Content cleaning and summarization,
- And web-based display of results.

It achieves the following benefits:

- Saves time for users by summarizing multiple articles.
- Improves efficiency in gathering information.
- Provides clean and relevant outputs without unnecessary text.

The system is modular, scalable, and can be extended in the future by:

- Adding more sophisticated summarization models,
- Introducing document ranking based on content quality,
- Or even providing source citations with each summary.