AI Web Analyzer - Digested Insights, Delivered Smartly.

COMP4801 Final Year Project Detailed Project Plan

Student: Fung Ho Sang (3035927386)
Project Supervisor: Prof. Yu Tao

Table of Contents

1. Background

The AI Web Analyzer is a final year project focused on personalized web content delivery through an GenAI-embedded chatbot. In today's fast-paced information environment, users face significant challenges in staying updated with relevant content due to the absence of standardized RSS feeds across many websites. This project addresses these challenges by integrating artificial intelligence to analyze and push relevant updates, enhancing users' ability to stay informed efficiently.

1.1 Current Challenges in Information Delivery

In today's digital landscape, users encounter significant challenges in accessing timely and relevant information. The lack of standardized RSS feeds across many websites complicates the process of staying updated, forcing users to manually check multiple sources. This fragmentation leads to inefficiencies and can result in missed opportunities to stay informed about important developments.

As highlighted in various literature [1], users frequently experience "information overload," making it increasingly difficult to sift through vast amounts of content to identify what is relevant. This emphasizes the necessity for automated systems that can simplify content delivery and alleviate the cognitive burden on users.

The AI Web Analyzer addresses these challenges by leveraging artificial intelligence to analyze user preferences and deliver personalized content updates. By automating the information retrieval process, this project aims to enhance users' ability to stay informed with minimal effort, significantly improving their overall experience in navigating the digital information landscape.

1.2 The Role of AI in Content Personalization

AI enhances content personalization by analyzing user preferences, enabling behavioral tailoring that delivers tailored experiences. Sodiya et al. [2] illustrate how AI-driven personalization boosts user engagement by presenting relevant information based on user behavior, significantly improving overall satisfaction.

Additionally, AI improves content retrieval speed, allowing for faster and more accurate access to relevant information. Research by Smith et al. [3] indicates that AI systems can analyze user behaviors to provide personalized recommendations, saving time and minimizing the effort needed to retrieve relevant content.

1.3 Chatbots as Information Gateways

Chatbots serve as efficient tools for spreading information, providing interactive platforms for users to receive updates and engage with content seamlessly. Chatbots offer advantages over traditional search methods, including simplified information retrieval process, providing quick and relevant responses, which reduces manual search time.

1.4 Purpose of Project

The AI Web Analyzer aims to revolutionize information delivery by providing a proactive solution that digests content from selected websites. This approach not only overcomes the limitations faced by users in finding specific updates but also acts as an example for future developments in personalized content delivery systems.

By leveraging AI and chatbots, the project promises to enhance how individuals interact with digital content, making it more accessible and tailored to their unique needs.

2. Objectives

- **Develop a SaaS Product**
    - Create a service, such as a Discord Bot, that uses AI to digest content of user-selected websites. This product will allow users to utilize the service within their existing platforms, offering a convenient and accessible tool for information retrieval.
    - Ensure that the service is scalable to handle numerous users and adaptable to various domains, from news to niche interest groups.
- **AI Digested Content Delivery**
    - Implement a GenAI embedded system that pushes daily updates to users based on their preferences. This involves developing algorithms that learn user interests over time, providing increasingly relevant content.
    - Incorporate user feedback mechanisms to refine and improve content recommendations continually.
- **User Engagement**:
    - Enhance user satisfaction by delivering personalized content and allowing users to ask follow-up questions. This involves integrating interactive elements within the chatbot to facilitate deeper engagement.
    - Develop features that allow users to customize their experience, such as setting notification preferences and choosing content categories.

3. Example Use Cases

3.1 Get the Newest Update from a Football Club

Objective:
Provide fans with the latest news, match updates, and player information from their favorite football clubs.

Implementation:
- Scrape official club websites and trusted sports news outlets for real-time updates.
- Use AI to filter and prioritize news based on user-defined interests, such as match results, transfer rumors, or player interviews.
- Deliver updates via a chatbot, allowing users to ask for specific information, such as upcoming matches or league standings.

Example:
- A user chooses their favorite club and sets preferences for the types of updates they want.
- The AI Web Analyzer sends notifications about players' news feed, upcoming matches, and post-match analyses.

3.2 Follow a Company's News Feed/Career Page

Objective:
Help job seekers stay informed about their target companies, aiding in interview preparation and awareness of new job openings.

Implementation:
- Monitor company websites, news feeds, and career pages for updates on corporate developments and job postings.
- Use AI to highlight news that could be relevant for interviews, such as new product launches or executive changes.
- Notify users of new job openings that match their skills and interests.

Example:
- A user targets a tech company and receives alerts about recent projects, partnerships, and job openings that align with their career goals.
- The system provides insights into company culture and strategic directions, helping the user prepare for interviews.

3.3 Get Alerts on Cheap Flight Tickets

Objective:
Assist travelers in finding affordable flights by providing timely alerts on price drops and special offers.

Implementation:
- Scrape airline websites and travel deal platforms for flight prices and promotions.
- Use algorithms to track price changes and identify patterns or trends in ticket pricing.
- Send alerts for flights that match user-defined criteria, such as destination, budget, and travel dates.

Example:
- A user sets up alerts for flights from Hong Kong to anywhere within a specific budget and time frame.
- The AI Web Analyzer notifies the user of a flash sale, allowing them to book at a reduced price.

3.4 Get Daily News Digest

Objective:
Provide users with a concise summary of daily news tailored to their interests.

Implementation:
- Aggregate news from multiple sources, including major news outlets and niche publications.
- Use natural language processing to summarize articles and identify key points.
- Deliver a personalized digest that covers the most relevant stories, categorized by user preferences.

Example:
- A user interested in technology and world politics receives a daily email or chatbot message summarizing top stories in these areas.
- The digest includes brief summaries and links to full articles for deeper exploration.

4. Methodology

4.1 Tech Stack

- **Web Scraping:**
  - Develop algorithms with Python library (Selenium) to collect real-time data from the user-specified websites for dynamic content rendering.
  - Test the reliability of the algorithm by testing with various websites.
- **AI Integration:**
  - Develop Node.js script to utilize Google's Gemini 1.5 Flash API, a Natural Language Processing (NLP) GenAI model, to analyze content.
  - By prompt engineering, extract relevant information and personalize the output according to the user's preference.
- **Notification System:**
  - Design a mechanism allowing users to select the websites or topic of interest that they want to receive information from.
  - Develop Node.js script to utilize Discord API as a medium for user interaction, including pushing notifications and answering user's follow-up questions.
  - Ensure system scalability and reliability for handling multiple users.

4.2 Ethical Consideration

This project will adhere to ethical standards in web scraping by complying with website terms of service and ensuring user data privacy. All website content will be scraped with the same visibility as a general public, and all user preferences are strictly anonymous to the GenAI model.

5. Project Schedule and Milestones

Phase	Period	Objectives
1	Aug 2024 - Oct 2024	Research on digested content delivery with GenAI Deliverables: - Detailed project plan - Project web page
2	Oct 2024 - Nov 2024	Design and develop a robust web scraping module to gather data from a web page efficiently. Prototype and test initial scraping scripts.
3	Nov 2024 - Dec 2024	Integrate AI models to process and analyze the collected data, ensuring relevance and accuracy.
4	Dec 2024 - Jan 2025	Implement a notification system that delivers updates via a chatbot, focusing on user-friendly interaction. Deliverables: - Preliminary implementation - First Presentation - Interim report
5	Jan 2025 - Mar 2025	Conduct testing with various use cases, such as pushing updates of the latest football news and job openings. Finalize all components and deploy the system.
6	Mar 2025 - Apr 2025	Final Deliverables: - Finalized and tested implementation - Final report - Final presentation

References

[1] J. Spira, "Information Overload: A Systematic Review of the Literature," Information Overload Research Group, 2019. [Online]. Available: https://www.researchgate.net/publication/265906917_Information_Overload_A_Systematic_Literature_Review.

[2] E. O. Sodiya, O. O. Amoo, U. J. Umoga, and A. Atadoga, "AI-driven personalization in web content delivery: A comparative study of user engagement in the USA and the UK," World Journal of Advanced Research and Reviews, vol. 21, no. 2, pp. 887–902, Feb. 2024. [Online]. Available: https://doi.org/10.30574/wjarr.2024.21.2.0502.

[3] B. Smith, A. Johnson, and R. Lee, "Exploring the Benefits of AI for Content Retrieval," Journal of Information Science, vol. 45, no. 1, pp. 1-10, 2023. [Online]. Available: https://www.researchgate.net/publication/378575962_Exploring_the_Benefits_of_AI_for_Content_Retrieval.