**Automated Web Scraping System for Alzheimer's Disease Research Articles from PubMed**

**1. Introduction**

Background:

Alzheimer's disease (AD) is the most common form of dementia, affecting millions worldwide. It is characterized by the progressive degeneration of brain cells, leading to severe cognitive impairment. Early detection and ongoing research are vital for understanding and managing this disease.

Objective:

The project aims to develop an automated web scraping system that enables users to retrieve specific information related to Alzheimer’s disease from PubMed articles using keywords. This system uses a one-time API key setup to streamline the data retrieval process, making it efficient and user-friendly.

**2. Data Sources**

PubMed:

PubMed, managed by the National Institutes of Health, is a critical resource for accessing biomedical literature. It offers extensive data on various health-related topics, including Alzheimer’s disease.

API Access:

The system uses a PubMed API key, obtained during the initial setup, to facilitate automated queries. This key is stored securely and used to authenticate requests, ensuring that users can access PubMed data without needing to manage API credentials repeatedly.

**3. Methodology**

Automated Search and Retrieval:

Step 1: Initial configuration involves setting up the PubMed API key and defining base URL parameters for article searches.

Step 2: Users input keywords related to Alzheimer’s disease, which are used to construct search queries.

Step 3: The system automatically fetches articles matching the keywords. Information such as titles, abstracts, publication dates, and authors is retrieved.

Step 4: Users can then select specific articles to obtain detailed information or further explore the topics of interest.

**Data Handling:**

The retrieved data is processed and presented in a user-friendly format, allowing for easy access and analysis.

**4. Challenges and Solutions**

API Rate Limits:

To address potential issues with PubMed’s API rate limits, the system includes error handling mechanisms that detect limit breaches and pause requests, resuming them at a safe interval.

**5. Future Work**

Expansion to Other Databases: Extend the system to scrape data from other medical research databases like Scopus or Web of Science.

Real-Time Data Updates:

Implement features to track and update data on new research publications automatically.

Enhanced Search Capabilities:

Develop advanced search algorithms to refine the results further and include semantic search capabilities.

**6. Conclusion**

This project provides a robust and efficient tool for researchers, healthcare professionals, and the public to access up-to-date information on Alzheimer’s disease. The automated system simplifies the data retrieval process, making relevant research readily accessible with minimal input.