

Project #1 for the Biomedical Information Retrieval Course

Name: Phan Ben

ID Student: P76127051

Project Overview:

The PubMed Document Retrieval project aims to create a user-friendly web application that allows users to upload PubMed XML files, search for articles using keywords, and view the search results with highlighted keyword occurrences. The project leverages Streamlit, XML parsing, and text highlighting techniques to achieve this goal.

Key Features:

1. **File Upload:** The application provides users with a sidebar for file management, allowing them to upload PubMed XML files. Users can upload multiple files simultaneously, simplifying the process of importing data.
2. **Search Functionality:** The heart of the application is its search functionality. Users can enter keywords in a text input field and choose whether the search is case-sensitive or not. The system then searches through the uploaded PubMed articles to find matches.
3. **Text Highlighting:** When a search is performed, the application highlights the matching keywords within the retrieved articles. This feature provides a visual aid for users to quickly identify relevant information.
4. **Detailed Article Information:** For each matching article, the application displays detailed information, including the PMID (PubMed ID), article title, journal title, ISSN (International Standard Serial Number), publication date, abstract, authors, and keywords.
5. **Statistics Display:** To provide additional insights, the application calculates and displays document statistics for each article, including the number of characters, words, and sentences.

User Experience:

The PubMed Document Retrieval application is designed with a user-friendly interface. Users can easily upload files, perform searches, and view search results with highlighted keywords. The clear and organized presentation of article information enhances the overall user experience.