

PubMed Article Summarization Web Application

Umema Ashar | 22i-2036 | BS Data Science

Overview

This documentation provides a comprehensive guide to developing a web application that summarizes PubMed articles. The application employs Flask as the backend framework and includes functionalities for user authentication, text preprocessing, extractive summarization, and file saving. This guide covers data exploration, model selection (for summarization), fine-tuning (optional), and web application development.

Setup Instructions

1. Clone repository:

```
git clone https://github.com/umema2004/pubmed_summarization  
cd pubmed-article-summarizer
```
2. Install Dependencies:

```
pip install -r requirements.txt
```
3. Run the Application:

```
python app.py
```

Model Selection and Fine-tuning

Extractive Summarization: Select sentences from the original text based on importance scores (e.g., word frequency, TF-IDF). Implement extractive summarization using NLTK for tokenization, stopwords removal, and sentence scoring.

Web Application Development

Framework:

- **Flask** is used for backend development due to its simplicity and flexibility in handling HTTP requests and responses.

Functionality:

- **User Authentication:** Implement login and logout functionalities using sessions and a simple CSV-based user database.
- **Text Preprocessing:** Clean and preprocess input PubMed articles to remove stopwords and non-alphanumeric characters.
- **Summarization:** Implement extractive summarization based on word frequency or integrate generative models for abstractive summarization (optional).
- **File Handling:** Allow users to input text directly or upload a file containing PubMed article data.
- **Output:** Display both the original article and its summarized version on the web interface. Optionally, generate an image based on the text input.

Optional Features (Advanced):

- **User Authentication:** Manage user sessions securely with Flask sessions and ensure credentials are stored and verified.
- **Customization:** Allow users to select the style or length of the summary (brief, detailed).
- **File Saving:** Enable users to save the original article and its summary as text files for future reference.