Word Extractor

Top N most frequent words.

MERN Stack and Python

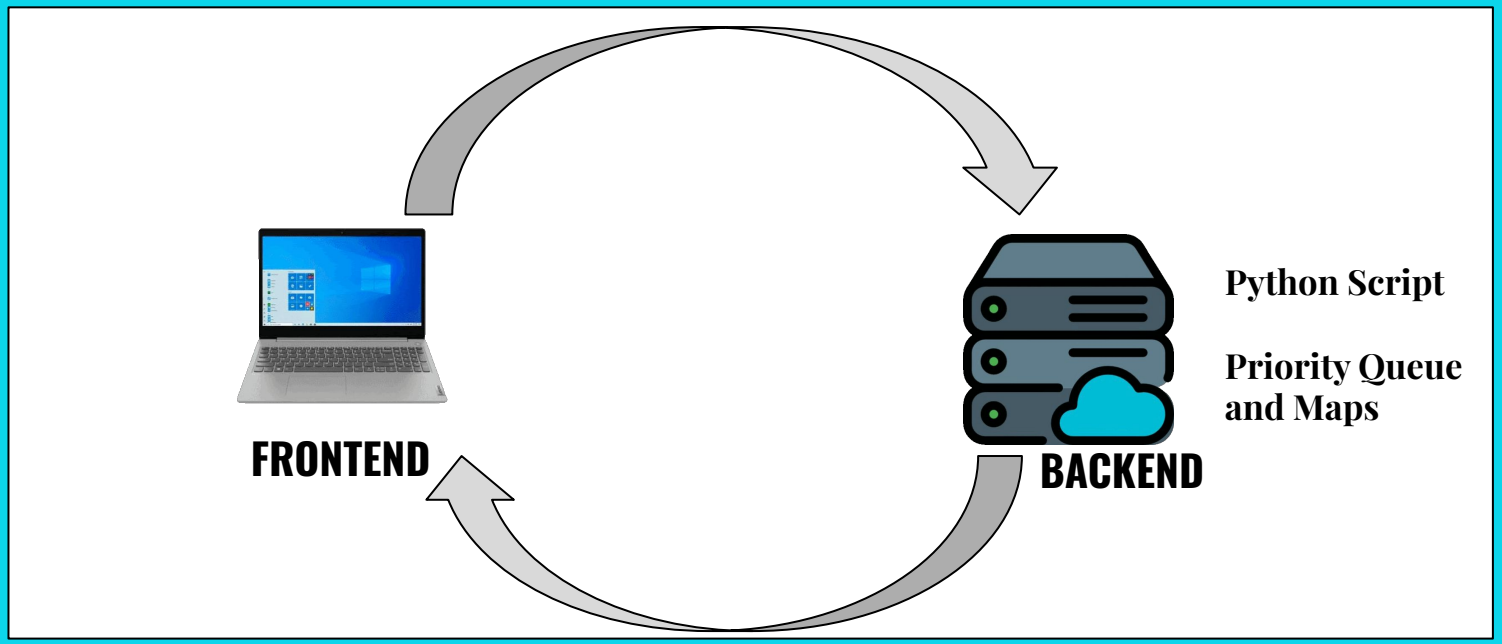
Tech Stack

- **Frontend: React.js**
- **Backend: Node.js, Express.js**
- **Python: Web scraping and word frequency analysis**
- **Data Structure: Priority Queue , Map**

Project Overflow

- 1. User inputs URL and top N value
- 2. Node.js server sends request to Python script
- 3. Python script scrapes, analyzes, and returns top N words
- 4. Frontend displays results

WorkFlow



Priority Queue in Word Frequency Analysis

Priority Queue Definition:

- Stores words by frequency for efficient access

Sorting:

- Highest frequency words appear first, ensuring top N words are easily retrieved.

Python Script - Code Overview

Modules: requests, BeautifulSoup, re

Steps:

- Fetch URL content
- Parse HTML and extract text
- Count word frequency
- Sort words by frequency

Future Enhancements

- 1. Add data visualizations for word frequencies
- 2. Optimize performance for large data
- 3. Store frequent analyses in MongoDB