**Search Engine Project**

This repository contains a simple **Search Engine** implemented for an **Information Retrieval** task. It includes components for web crawling, text preprocessing, stemming, and a ranking mechanism using the **Vector Space Model**.

**Features**

* **Web Crawler**: Fetches web pages for indexing.
* **Text Preprocessing**: Tokenization, stopword removal, and case normalization.
* **Stemming**: Reduces words to their root form.
* **Vector Space Model**: Calculates term weights and ranks relevant documents based on user queries.

**Requirements**

* Python 3.x
* Required Libraries: numpy, requests, beautifulsoup4

Install dependencies with:

pip install numpy requests beautifulsoup4

**Setup and Usage**

1. **Clone the repository**:
2. git clone https://github.com/SudhaChandrikaY/SearchEngine.git
3. cd SearchEngine
4. **Run the Search Engine**:  
   Execute the application with:
5. python app.py
6. **Search**:
   * Enter a search query when prompted.
   * The program will output ranked results based on relevance.

**Project Structure**

* **WebCrawler.py**: Crawls web pages and extracts text.
* **PreProcessor.py**: Handles text cleaning and stopword removal.
* **Stemmer.py**: Implements stemming algorithms.
* **VectorSpaceModel.py**: Processes term weights and query rankings.
* **stopwords.txt**: A list of stopwords for filtering.
* **app.py**: Main script to run the search engine.

**License**

This project is licensed under the **MIT License**.

**Author**: [SudhaChandrikaY](https://github.com/SudhaChandrikaY)

This format is simple, professional, and provides all necessary details concisely.