

News Article Search Engine (CS516 Term Project)

- Author: Raheel Riaz
- GitHub Repository: <https://github.com/mscs24035/IRCourseProject>

Motivation & Problem Statement

- Need fast and relevant news retrieval
- Demonstrate IR concepts: indexing, ranking, boolean search
- Implement a functional news search engine

System Architecture

- Dataset → Preprocessing → TF-IDF Indexing
- Boolean AND search
- Ranked retrieval (TF-IDF + Cosine Similarity)
- Flask Web Interface for searching

Implementation Details

- TF-IDF vectorizer using scikit-learn
- Preprocessing: tokenization + stopwords removal
- Search methods: Boolean AND + Ranked search
- Flask UI for real-time querying

How to Run the Demo

- 1. `python src/indexer.py` (build index)
- 2. `python src/app.py` (launch UI)
- Open: `http://127.0.0.1:5000`
- Try queries: “AI model”, “climate summit”, “local elections”

Results & Features

- Functional ranked retrieval
- Boolean search supported
- Lightweight clean UI
- Easily extendable codebase

Future Enhancements

- BM25 ranking
- Larger datasets (BBC, Kaggle)
- Stemming & Lemmatization
- Optimized inverted index
- Improved UI

Thank You

- GitHub Repository:
- <https://github.com/mscs24035/IRCourseProject>

