NLP and IR - Natural Language Processing and Information Retrieval
자연어처리 및 정보검색

- **Billboard Ranking Source**: Billboard Hot 100
  - **Date of Data**: 24.03.25
  - **GitHub Repository**: NLP-and-IR
  -

Project Overview

For this project, I decided to collect the **top 100 Billboard songs** and use their lyrics along with additional metadata like **artist name, song title, etc.**.
To obtain the lyrics, I used Lyricsify and converted them into .txt files for easier processing.
The **Information Retrieval (IR) system** is built in **Python**, as it provides a simple and effective approach for this assignment.

Implementation Details

✅ **Boolean Search**: Supports **AND, OR, NOT** operators (currently works when not mixed).
✅ **Vector Space Model**: Uses **TF-IDF weighting** for ranked search results.
✅ **UI**: A basic **Tkinter** interface to select search type, enter queries, and display results.
✅ **Future Improvements**:
- Boolean search is **not yet case-sensitive** or fully syntax-aware (planned for future updates).
- UI is currently simple, but I plan to enhance it with additional dependencies like ttkbootstrap (is that allowed?).
-

How to Run the Project?

🎵 **Simply start main.py** – that's it! 😃
All project files, including the dataset and code, are available on **GitHub**:
🔗 NLP-and-IR Repository