Traditional NLP Project

Objective:

Build an NLP-powered movie information system using the IMDb Top 100 movies. This project will guide you through data scraping, fuzzy matching, and similarity-based recommendations.

Task 1: Build the IMDb Top 100 Movie Dataset

Description:
Create a dataset containing detailed information for the top 100 movies listed on IMDb.

Requirements:
Visit the IMDb Top 100 Movies page: <u>IMDb Top 250 movies</u>
For each movie, collect the following:
- ✓ Title
- ✓ Duration
- ✓ Year of Release
- ✓ IMDb Rating
- ✓ Description (Plot Summary)
- ✓

Output:
Save the data in a CSV file named top_100_movies.csv

Notes:
- ✓ First, extract the links to individual movie detail pages.
- ✓ Base URL to construct movie links:
      base_url = "https://www.imdb.com"

Task 2: Query-Based Movie Lookup

Description:
Allow users to input a movie title and retrieve its full information from the dataset.

Requirements:
- ✓ The system should accept a movie title input from the user.
- ✓ Use fuzzy string matching (e.g., Levenshtein distance or fuzzywuzzy library) to correct for spelling errors in the user input.
- ✓ Display all available information (from the CSV) about the matched movie.

Task 3: Movie Recommendation Using Description Similarity

Description:
Using the description of the movie retrieved in Task 2, recommend 5 similar movies based on content similarity.

Requirements:
- ✓ Use text similarity techniques
- ✓ Rank and display the top 5 most similar movies based on description content.