



Rhombus AI

Web Application for Regex Pattern Matching and Replacement

Objective:

You are tasked with creating a web application using Django and React that allows users to upload CSV or Excel files, identify patterns in text columns using natural language input, and replace the matched patterns.

Requirements:

1. Backend Development with Django:

- Set up a Django project for data processing.
- Implement models, views, and URLs to handle data processing logic and user interactions.
- Create an API endpoint to receive natural language input, convert it to regex using an LLM, and handle the replacement operations.

2. Frontend Development with React:

- Develop a user-friendly interface allowing users to upload CSV/Excel files.
- Provide input fields for users to describe the pattern they want to match in natural language and specify the replacement value.
- Display the processed data with the applied replacements in the text columns.

3. LLM Integration:

- Use a Large Language Model (LLM) to convert natural language input into a regex pattern.
- Ensure the LLM can handle various natural language descriptions accurately.

- [OPTIONAL] We encourage you to showcase your creativity by using LLM for two data transformations of your choice.

Task Description:

1. File Upload:

- Users should be able to upload CSV or Excel files.
- The uploaded file should be processed and displayed in a tabular format.

2. Pattern Matching and Replacement:

- Users describe the pattern they want to find in natural language (e.g., "find email addresses").
- The application uses an LLM to convert this description into a regex pattern.
- Users specify the replacement value for the matched pattern.
- The application replaces the matched patterns in the text columns and displays the updated data.

Example Scenario:

Sample Input Data:

ID	Name	Email
1	John Doe	john.doe@example.com
2	Jane Smith	jane_smith@domain.com
3	Alice Brown	alice.brown@website.org

User Input:

- **Natural Language:** "Find email addresses in the Email column and replace them with 'REDACTED'."

LLM Output:

- **Regex Pattern:** `\b[A-Za-z0-9._%+-]+\@[A-Za-z0-9.-]+\.[A-Za-z]{2,7}\b`

Replacement:

- **Replacement Value:** 'REDACTED'

Processed Output Data:

ID	Name	Email
1	John Doe	REDACTED
2	Jane Smith	REDACTED
3	Alice Brown	REDACTED

Development Guidelines:

- Write clean, maintainable, and well-documented code.
- Include reasonable error handling and validations in both backend and frontend.
- Good software design is very crucial.
- **[OPTIONAL]** Support for large files.

Deliverables:

1. Source Code:

- Submit the complete source code via a GitHub repository.

2. README File:

- Provide a detailed README.md file that includes:
 - Instructions for setting up and running the application.
 - Additional notes or comments about the project.

3. Demo Video:

- Submit a short video demonstrating the web app in action, embedded in the GitHub README.md file.