Text Similarity API Using Django

Introduction

Django is a powerful and popular web framework for building web applications in Python. It emphasizes rapid development, a clean and pragmatic design, and follows the "Don't Repeat Yourself (DRY)" and "Convention Over Configuration" principles. In this project, Django is utilized to create a RESTful API for calculating text similarity. For training this model I've used dataset which contains id and two other columns which contains text paragraphs.

Libraries and Tools

Python: The primary programming language used for this project.

Django: The web framework for building the API.

NLTK: A natural language processing library for text preprocessing.

Gensim: A library for training Word2Vec models.

Postman: A popular tool for testing APIs and sending various types of HTTP requests.

Creating the Django Project

To start a Django project, use the following command:

- ➔ django-admin startproject text_similarity_project
- ➔ cd text_similarity_project

API Endpoints

1. Train Word2Vec Model

Endpoint: `/train_word2vec/`

HTTP Method: POST

Description: Trains the Word2Vec model using provided text data.

Example Input (Request Body)

```
{
    "text1": "First text paragraph.",
    "text2": "Second text paragraph."
}
```

Example Output (Response)

```
{
    "message": "Word2Vec model trained successfully."
}
```

2. Calculate Similarity

Endpoint: `/calculate_similarity/`

HTTP Method: POST

Description: Calculates the similarity between two text paragraphs using the trained Word2Vec model.

Example Input (Request Body)

```
{
    "text1": "First text paragraph.",
    "text2": "Second text paragraph."
}
```

Example Output (Response)

```
{
    "similarity_score": 0.85
}
```

Running the API

1. Navigate to the project directory using the terminal.

2. Start the Django development server:

➔ python manage.py runserver