

PLAGIARISM ANALYSER

Abstract:

Plagiarism Analyzer is a software tool designed to detect and analyze instances of plagiarism in written documents. This documentation outlines the development, implementation, and functionality of the Plagiarism Analyzer system. The tool utilizes advanced algorithms and techniques to compare texts, identify similarities, and generate detailed reports to aid in the detection and prevention of plagiarism. This documentation provides an in-depth overview of the system architecture, features, and usage instructions.

Introduction:

Plagiarism is a serious issue in academia, journalism, and various other fields where originality and integrity are paramount. With the increasing availability of digital content, detecting plagiarism has become more challenging yet crucial. Plagiarism Analyzer aims to address this challenge by providing a reliable and efficient tool for identifying plagiarized content.

Literature Survey:

Prior research in the field of plagiarism detection has led to the development of various techniques and algorithms. These include methods based on text similarity metrics, natural language processing (NLP), machine learning, and deep learning. Existing plagiarism detection systems vary in their approaches, from simple string matching to sophisticated semantic analysis. By reviewing the literature, we aim to identify the strengths and weaknesses of different approaches to inform the design and implementation of Plagiarism Analyzer.

Models:

Plagiarism Analyzer employs several models and algorithms to analyze textual content and detect instances of plagiarism. These models include:

- a. Text Preprocessing Model:** This model prepares the input text for analysis by removing noise, such as punctuation and stop words, and normalizing the text for consistency.
- b. Similarity Detection Model:** This model compares the preprocessed texts using various similarity metrics, such as cosine similarity, Jaccard similarity, or Levenshtein distance, to identify similarities between documents.
- c. Machine Learning Model:** Plagiarism Analyzer may utilize machine learning algorithms, such as Support Vector Machines (SVM), Naive Bayes, or neural networks, to classify texts as plagiarized or non-plagiarized based on features extracted from the comparison results.

Methodology:

The methodology employed by Plagiarism Analyzer involves the following steps:

- a. **Input Text Retrieval:** Obtain the texts to be analyzed from the user or from external sources such as files or databases.
- b. **Text Preprocessing:** Clean and preprocess the input texts to remove noise and prepare them for analysis.
- c. **Similarity Comparison:** Compare the preprocessed texts using appropriate similarity metrics to identify similarities and potential instances of plagiarism.
- d. **Feature Extraction:** Extract relevant features from the comparison results, such as similarity scores or patterns, for further analysis.
- e. **Plagiarism Classification:** Utilize machine learning algorithms to classify the texts as plagiarized or non-plagiarized based on the extracted features.
- f. **Report Generation:** Generate detailed reports outlining the detected instances of plagiarism, including the similarity scores, matched passages, and recommendations for further action.

By following this methodology, Plagiarism Analyzer provides users with a comprehensive tool for detecting and analyzing plagiarism in written documents.

This documentation serves as a guide for users and developers interested in understanding the functionality and implementation of the Plagiarism Analyzer system.

Source Code:

```
from difflib import SequenceMatcher

with open('charminar1.txt') as one_file, open('charminar2.txt') as two_file:
    data_file1 = one_file.read()
    data_file2 = two_file.read()
    matches = SequenceMatcher(None, data_file1, data_file2).ratio()
    print(f"The plagiarized content is {matches * 100:.3f}%")
```

Result:

Charminar, situated in Hyderabad, India, is a renowned monument celebrated for its grandeur. The structure boasts four majestic minarets, hence its name "Charminar," which translates to "Four Towers" in Urdu and Persian. Constructed in 1591 by Sultan Muhammad Quli Qutb Shah, ruler of the Qutb Shahi dynasty, this architectural marvel is crafted from granite, limestone, and mortar. Nestled in the heart of the old city, each minaret comprises four stories with spiral staircases leading to the top, where a mosque once welcomed worshippers.

Charminar's architectural style reflects a fusion of Indian, Persian, and Islamic influences, characterized by intricate carvings and embellishments adorning its walls. Beyond its structural beauty, Charminar holds cultural significance, serving as a symbol of Hyderabad's rich heritage. Surrounded by a vibrant market area, visitors flock to Charminar to immerse themselves in its history, explore its intricate design, and indulge in the bustling atmosphere of the surrounding marketplace.

Charminar is a famous monument located in Hyderabad, India. It is a big and beautiful structure with four towering minarets, which is why it's called "Charminar," meaning "Four Towers" in Urdu and Persian. The monument was built in 1591 by Sultan Muhammad Quli Qutb Shah, who was the ruler of the Qutb Shahi dynasty.

The Charminar is made of granite, limestone, and mortar, and it stands in the heart of the old city. Each of its four minarets has four stories with spiral staircases inside. At the top, there is a mosque where people used to pray.

One of the interesting things about Charminar is its design. It blends elements of Indian, Persian, and Islamic architecture. There are also intricate carvings and decorations on its walls, adding to its beauty.

Charminar is not just a monument; it's also a symbol of Hyderabad. It's surrounded by a bustling market area where you can find all sorts of things, from jewelry and clothing to delicious street food. Many people visit Charminar to enjoy its beauty, learn about its history, and experience the vibrant atmosphere of the surrounding area.

The plagiarized content is 21.536%