

Dylan Boyle | OOSD

Assignment 2024 - Similarity

Search with Word

Embeddings

Words Embeddings Text Interface

This document provides an overview of the various components and functionalities available in the Words Embeddings Text Interface project. The project allows for loading and interacting with word embeddings files, performing vector operations, and finding similar or dissimilar words using various similarity metrics

Console

ConsoleColor An enumeration for terminal color codes to enable colored output in the console.

ConsoleLoadingMeter Static methods for displaying a text-based progress meter in the console to indicate the progress of loading or processing tasks.

ConsolePrint Static methods for consistent message printing to the console, ensuring uniform output format across the application.

Embeddings

WordsEmbeddingsSearch This class is responsible for loading and interacting with a words-embeddings file. It includes functionalities such as:

- CSV format detection to determine the delimiter used in the embeddings file.
- Word vector operations including search, add, subtract, multiply, and divide.
- Similarity metrics to find similar or dissimilar words based on the vector representations.

Menu

ConfigureMenuObject An enumeration for settings menu options with auto-generated shortcuts. It provides an easy way to navigate and configure application settings.

ConfigureOptionsMenu This class stores user preferences with persistent values and includes default settings. It allows users to configure various aspects of the application, including:

1. **Load New Word-Embeddings File** Load a new file containing word embeddings.
2. **Load New Data-Output File** Specify a new output file for storing results.
3. **Number of Similarities to Find** Set the number of similar words to find ($0 < n \leq \text{\#words}$).
4. **Select Similarity Metric to Use** Choose the similarity metric for finding similar/dissimilar words.
5. **Toggle Similarity Score in Output** Enable or disable the inclusion of similarity scores in the output.
6. **Awitch between Append/Overwrite Modes Data-Output File** Choose whether to append to or overwrite the output file.
7. **Clear the Output File** Clear the contents of the output file.
8. **Reset Configure Options to Defaults** Reset all configuration options to their default values.
9. **Print Current Configuration Options** Print the current configuration settings.
- q. **Close Configure Options** Exit the configuration menu.

MainMenu Handles user interactions and application launches. It prompts for a words-embeddings file if not loaded and provides the following options:

1. **Specify Embedding File** Specify the file containing word embeddings.
2. **Specify an Output File (default: ./output.txt)** Specify the file to store the output.
3. **Enter a Word or Text for 'Similiar Words'** Find the top matching words for a given input.
4. **Enter a Word or Text for 'Dissimilar Words'** Find the least matching words for a given input.
5. **Run 'Text Calculator'** Perform vector operations and find top matching words.
6. **Configure Options** Customize application preferences.
- q. **Quit** Exit the application.

MainMenuObject An enumeration for main menu options with auto-generated shortcuts, making it easier to navigate the main menu.

TextCalculatorMenu Provides options for performing vector operations using words' embeddings:

1. **Add** Add two word vectors.
2. **Subtract** Subtract one word vector from another.
3. **Multiply** Multiply two word vectors.
4. **Divide** Divide one word vector by another.

TextCalculatorMenuItem An enumeration for text-calculator-menu options, allowing users to select vector operations easily.

VectorSimilarityMetricsObject Turns the VectorSimilarityMetrics utility class into a menu, providing options for different similarity metrics.

Utilities

Array Static methods to find minimum or maximum values in an array, useful for determining the most similar or dissimilar words.

VectorOperations Static methods for vector operations like add, subtract, dot product, etc., providing the core functionalities for manipulating word vectors.

VectorSimilarityMetrics An enumeration with an abstract calculate method, representing various similarity metrics:

1. **Dot Product** Calculates the dot product between two vectors.
2. **Euclidean Distance (No Square Root)** Calculates the Euclidean distance without taking the square root.
3. **Euclidean Distance** Calculates the Euclidean distance between two vectors.
4. **Cosine Similarity** Calculates the cosine similarity between two vectors.