Similarity Search

@Nathan Buyrchiyev

@version Java 17

Description

This is a console-based application that can parse a user-defined .txt file containing words and their corresponding vector values. It allows users to choose a word from the embedding file they have chosen and receive a list of the most similar words to the word they have chosen according to its vector values.

To Run

From console at .jar file directory:

java -cp ./dsa.jar ie.atu.sw.Runner

Navigate through the menu options and select your options as desired.

If an embedding file is not chosen, you will not be able to search for a word.

Please note that if an output file directory is not chosen, "./out.txt" will be chosen as default.

Features

- Specify the text file to be embedded.
- Specify the directory of the output file to be created.
- Search for a word in the embedded file.
- Choose how many similar words you want printed on to the output file.
- An output file containing the most similar words to the input word.
- An option to choose between having similar words listed from most similar least similar and vice versa.