Similarity Search with Word Embeddings

Completed by student Andrii Skvortsov

@java version 21.0.2.0

Description: This Java application performs a similarity search using word embeddings to find the most semantically similar words or phrases. It processes word embeddings from a specified file and outputs the top matches based on similarity scores to a user-defined output file.

Main Features:

1. **Parse**:
   The word amount to be processed is not hardcoded and can be other than the default 59,602 words. The embedding number has to be strictly 50 numbers for each word.
2. **Search**:
   Utilizes cosine distance to measure the similarity between the input vector and the word embeddings. If one or more words are not found in the embeddings file, the user will be notified and the calculation will be performed with only matched words. The application computes an average vector for all matched words.
3. **Output**:
   Outputs the top `n` most similar words and their similarity scores to a specified output file.
4. **Menu Interface**:
   The command-line interface allows users to:
   - Specify the path for the word embeddings file (default: `./word-embeddings.txt`).
   - Specify the path for the output file (default: `./out.txt`).
   - Enter a word or a short sentence for similarity comparison. Configure the number of top matches (`n`) to display.
   - Perform the similarity search and output the results.
5. **Colored Console**:
   Enhances the user experience with ANSI escape codes for colored text output.

Instructions:

1. Place the `word-embeddings.txt` file in the project directory.
2. Run the application from the project directory.
3. Follow the menu prompts to configure file paths, input words, and perform the search.
4. View the output file for the list of similar words and similarity scores.