

Word Embedding Searcher

Author: Alaeddine Jmila

Version: Java 21

Description

This is a console-based application designed to return a number of matching results for a search term or phrase using Word Embedding. It utilises a word embeddings file generated by neural networks and outputs a text file with the results, well-formatted and displaying their match scores.

How to Run

1. From the console, navigate to the directory containing the `src` directory and run the following commands:
2. `javac ie/atu/sw/*.java` (to compile to bytecode)
3. `java ie.atu.sw.Runner` (to run the application)
4. Follow the console prompts to configure your files and settings.
5. If you do not specify the output file, default settings will be used. If the embedding file is not found, you will be prompted to provide its path.

Features

1. **Specify Embedding File**
 - Users can specify the file containing word embeddings to be used in the similarity search.
2. **Specify an Output File**
 - Users can specify the output file where results will be saved.
 - Default output file is set to `./out.txt`.
3. **Enter a Word or a Sentence**
 - Users can input a word or a sentence for similarity search.
4. **Configure Options**
 - Specify the Number of Results to Return (default: 10)
 - Specify the Maximum Number of Words to Process at Once (default: 10)
 - Choose Search Mode. Options: Whole sentence (A, default) / Individual words (B)
 - Return Unmatched Results. Options: Yes (default) / No (recommended if "individual words" mode is selected)

Additional Features Based on Context

- **Handling Long Sentences**
 - If the input sentence exceeds the maximum allowed number of words:
 - Option to continue processing the allowed number of words.
 - Option to enter a new sentence.
 - Option to increase the maximum number of words to process.