**README for Word Clustering with Virtual Threads Project**

**Project Overview:**

The Word Clustering with Virtual Threads project is a Java application designed to process word embeddings using the k-means clustering algorithm. It provides a menu-driven interface that allows users to input the path to the embeddings file, specify the number of clusters, and search for a specific word within the dataset. The project uses virtual threads for efficient processing and clustering of large datasets.

**Main Features:**

* **Menu-driven UI:** The application allows users to interact through a simple console-based interface.
* **Word Embedding Clustering:** Uses the k-means algorithm to group similar words based on their embeddings.
* **Search Functionality:** Allows users to input a search term and find the closest matching words in the dataset.
* **Virtual Threads for Efficiency:** The program uses Java's virtual threads for concurrent processing of word embeddings, improving performance on large datasets.

**How to Run the Application:**

1. **Requirements:**
   * Java Development Kit (JDK) 17 or later installed.
   * Command line or terminal access.
2. **Running the Program:**
   * Navigate to the directory where your JAR file (embeddings.jar) is located.
   * Use the following command to run the program:

* java -cp ./embeddings.jar ie.atu.sw.Runner
  + The application will prompt you for the file path of the word embeddings and the number of clusters. You can also search for specific words in the dataset.

1. **Inputs:**
   * **Embeddings file path:** The file that contains word embeddings data.

* **ie:** ./src/word-embeddings.txt (for my testing)
  + **Number of clusters:** The number of clusters to form during the k-means clustering.
  + **Search term:** The word you wish to search for.

1. **Output:**
   * Displays the closest matching words for the search term, based on the word embeddings.

**Directory Structure:**

The directory structure of the project is as follows:

<G00400530>.zip

|

|-embeddings.jar # The executable JAR file containing the compiled code

|-src # The source code folder containing the Java classes

|-README.pdf # PDF file with the project description

|-design.png # UML class diagram in PNG format

|-docs # Folder containing the Javadoc

|\_ index.html # Javadoc for the project

**Known Issues:**

* The program may experience slower performance with extremely large datasets.
* The search functionality currently only returns a set number of closest words (5).