

## **What is the application ?**

This application focuses on probabilities of a sequence of letters or words. The probabilities are important for augmentative and alternative communications systems and can be used for word prediction and word suggestions. The program uses the language model, n-grams. N is the length of the sequence of characters. The program uses a CLI and will ask the user to enter the number of n-grams required.

Here is a N-Gram CSV frequency table generated by the program:

N-Gram	
5-Gram	Frequency
month	139
rance	236
openl	231
anion	507
tfash	188
alley	499
troub	345
arkly	69
eenwa	93
rickh	149
sails	345
tfull	337
tract	423
ntyse	181
atter	394
leari	205
ences	176
ersto	175

The program asks the user to input a directory which reads the text and outputs the n-gram occurrences into a specified format. The program uses a text file to read and outputs a CSV file. The program was tested with a CSV file output and text file input.

## **What Happens When Application is Run ?**

The application should display the following options and the user can select the option they wish to select to make the CSV output

### **Specify Text File Directory**

The text file folder should be in the directory but can be accessed in from most places. To get the file path in Eclipse you should have the text file folder in the project directory and right click on the text file -> properties -> copy the location field and input the copied file path name into application. The program will find the input text file.

### **(2) Specify n-Gram Size**

When the user has to pick the n-gram size the user can pick the number of n-grams needed, 1-5 is recommended as the output could be large and rearrangement might be required.

### **(3) Specify Output File**

The user can pick a file name, CSV is recommended.

### **(4) Build n-Grams**

The program will output the file the user inputted option from the CLI menu driven application.

### **(5) Quit**

The program should exit.

## **What Features Does it have ?**

The program is a menu driven application which runs in the CLI. The program was built using Eclipse and Java SE 17 compiler. The program is able to take text input and output a file. The file will include the n-gram size the user has chosen. The file will have two columns N-Gram and Frequency and can output the results. The program is expected to store the results in an object 2D array and output the file to CSV. The program can be run from a jar file to access the program alternatively.