

N-Gram Frequency Builder

N-Gram Frequency Builder is a Java N-Gram Frequency Builder to get the N-Gram of characters found in a piece of text which is stored in a .txt file.

Installation:

Please run **Java -Xmx1G -cp ./ngrammer.jar ie.gmit.sw.Runner** in Command Prompt whilst in the folder where the .JAR file is located.

Usage:

Option 1: Choose File Dir of your .txt File.

E.g., C:\DSA_Project\DSA_Project\bin\books\BibleGod.txt

Option 2: Select N-Gram size.

E.g., 1 N-Gram will count how many single letters there are and so on.

Option 3: Choose the output file name this will automatically create the file into .CSV.

Option 4: This will execute N-Gram Calculation and output file where .JAR file is located.

Option 5: Exit application.

Features:

- All Menu items are in the menu class.
- User Needs to enter valid items into all fields before N-gram is calculated.
- All characters other than the alphabet from A-Z are disregarded.
- White Space is disregarded.
- Every N-Gram is outputted to lowercase.
- All Big – O Notation is commented above Each loop.
- User does not need to enter .CSV when entering output Filename.
- Automatically sorts data into 2 rows within .CSV File.
- Terminal Progress meter is built-in for visual effect.
- Books folder will be included with various books.
- Paths and environmental variables are not hard-coded.

Output Example: (N-Gram of 1 using BibleGod.txt)

	A	B
1	a	344443
2	b	62285
3	c	79169
4	d	199583
5	e	508476
6	f	104238
7	g	71833
8	h	337810

Patril screenshot of output.csv file.