README

This application returns the guess of the next word!

There is a box where you input and a box that will return the guess.

### Dataset

* 10 samples of 100,000 lines of were taken from each dataset(Twitter,News,Blogs)
* These samples were procesed creating 1-gram, 2-gram, 3-gram and 4-gram
* The top 100,000 n-grams from each where brought forward to be used in the application
* Using this approach reduced the dataset from several GB of data to under 10 MB of uncompressed data
* Processing of this data took several hours
* This enabled the application to search and perform queries quickly