**Names:**

Samar SalahElDin Ghoneim 20188029

Toka Mohamed ElDesoky 20188013

Somaya Yasser Galal 20188065

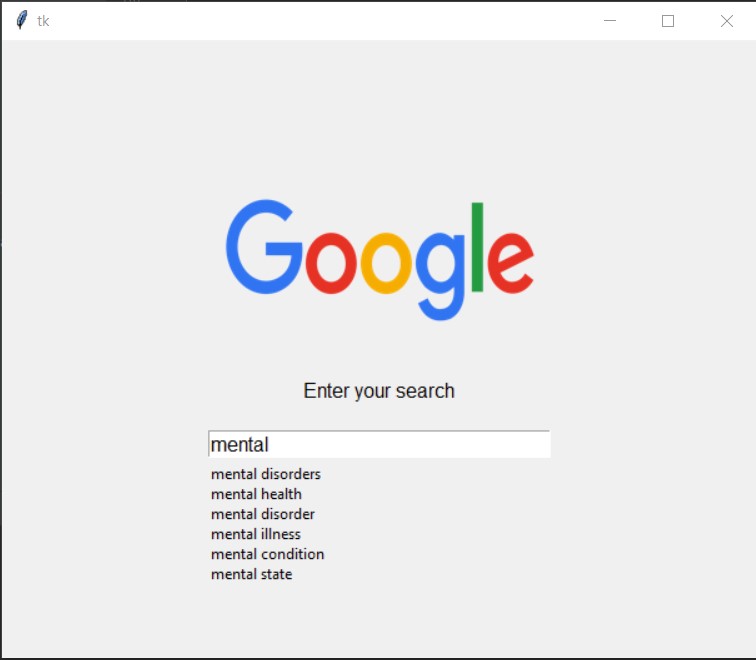
**Documentation:**

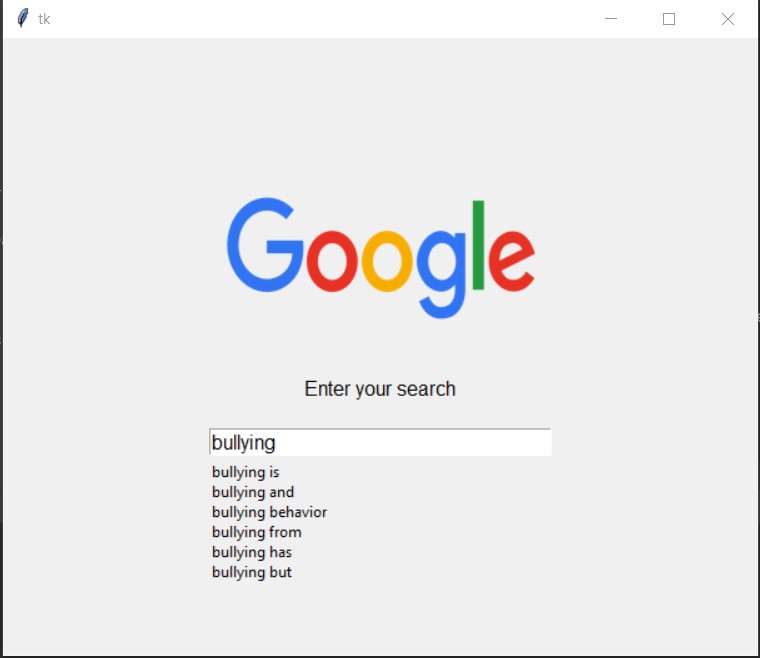
We gathered our corpus from a mental health book which was an encyclopedia about mental health; “Mental Health and Mental Disorders\_ An Encyclopedia of Conditions, Treatments, and Well-Being”. It had over 700 thousand words. We cleaned the data by removing the first few pages, images and the last few pages. We also removed links and unnecessary elements. We have also removed sections to remove some of the data because it was too big to run as it took a long time. After all of that, we ended up with over 200 thousand words.

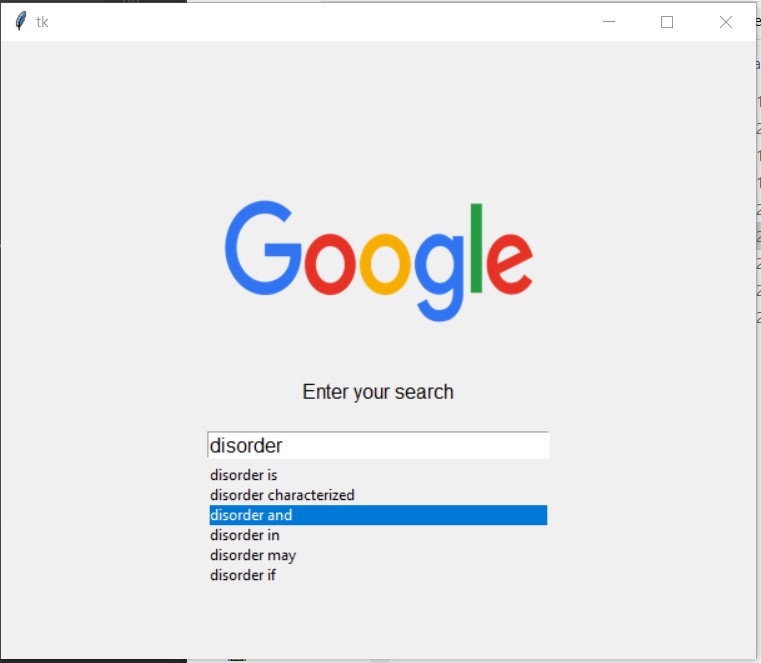
**Probability table:**

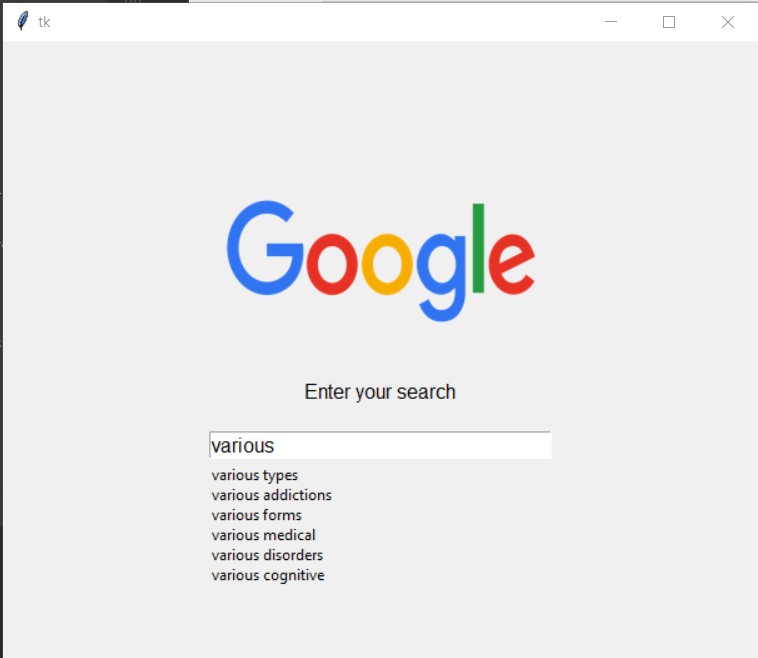
Attached the output in a csv file that has been generated by a python code. It shows the probability that a word (in row) followed a certain word (in column) and the relation of all unique words together (probability wise).

**Screenshots:**

****

****

****

****