Grader Output

Assignment4

100.00 / 100.00 points earned
5 / 5 autograded cells passed

Graded Cells

Cell 7 (cell-f46918d88bcfc44c

Passed | 40.00 / 40.00 points
View feedback

Cell 8 (cell-592d2f3552e2059

Passed | 20.00 / 20.00 points
View feedback

Cell 8 (cell-8a8072c81e421ae

Passed | 0 / 0 points
View feedback

Cell 11 (cell-c43547e8c520ff88

Passed | 40.00 / 40.00 points
View feedback

Assignment 4

In this assignment, we will explore countmin sketches and bloom filters. We will use two text files `great-gatsby-fitzgerald.txt` and `war-and-peace-tolstoy.txt` to load up the text of two famous novels courtesy of Project Guttenberg.

We will explore two tasks:

- Counting the frequency of words of length 5 or more in both novels using a count-min sketch
- Using a bloom filter to approximately count how many words in the War and Peace novel already appears in the Great Gatsby.

Cell 11 (cell-be6d976534e0936

Passed | 0 / 0 points

View feedback

Step 1: Making a Universal Hash Family (Already Done For You)

We will use a family of hash function that first starts by (a) generating a random prime number p (we will use the Miller-Rabin primality test for this purpopse); (b) generating random numbers a, b between 2 and p-1.

The hash function $h_{a,b,p}(n) = (an + b) \mod p$.

Note that this function will be between 0 and p-1. We will need to also make sure to take the hash value modulo m where m is the size of the hashtable.

To hash strings, we will first use python's inbuilt hash function and then use $h_{a,b,p}$on the result.

As a first step, we will generate a random prime number.

(A) Generate Random Prime Numbers

In [1]:
```
# Python3 program
Miller-Rabin rand
omized primality
test
# Copied from gee
ksforgeeks: http
s://www.geeksforg
```