

## Lab 3

Toward natural language processing:

- we need some base-level statistics practice
- everyone has COMP 251 prerequisite
  - you have a basic understanding of hash tables as a data structure
- the collisions in a hash table have an extremely rough approximation for conflating two concepts in a latent feature space, even though the hash table deals with discrete inputs whereas a latent space deals with continuous
- considering the probabilities of collisions, although not an NLP topic, gives a good intermediary exercise between COMP 251 and NLP

You are fine to work together with others. The tasks to complete are numbered below.

- 1) Design a function that produces a pseudorandom integer given an input integer in the domain  $[0, 99]$ , i.e.: a hash function. If you use a reference for the design, cite the URL.
- 2) Create your own hash table data structure that uses your pseudorandom function as a hash to store input strings. You will have to write another function that converts input strings into an integer to input into your hash function.

For the hash table data structure there should be the functionality to:

- create a table that is empty
  - insert new values into the table at a location based on your hash function
  - deal with collisions, either by probing, or use of linked-lists
- 3) If you choose to work with another person, write a brief comparison (up to three in a group) to describe the differences between the design of your pseudorandom hash function and the design of your hash table. Either if you work on your own or in a group, give some brief description of how often there is a collision in your hash table.

This lab is out of 5 marks total. There are 2 marks for each of Question 1 and 2, and 1 mark for Question 3. Submit a `lab3.hs` file to Blackboard in the Assignments and Tests section under Lab 3 submission. Deadline for submission is Saturday, Oct 22, before 11:59 pm.