Final Project CS2223 D-Term 2020

**Open Hashing (Separate Chaining)**

1. Implementation

Develop an Open Hashing (Separate Chaining) implementation (or find and document one).

2. Testing

Test your implementation on some small text file.

Test your implementation on “The Elephant’s Child”?

Test your implementation on an English language dictionary?

(I may provide you with one someone recently sent me.)

Test your implementation on “Moby-Dick”.

3. Comparison with another method

Compare and contrast your implementation with/on/against the Closed Hashing system you developed for HW5.

Does it matter whether the tables have the same size?

How big should they be?

What are the analogues for Open Hashing to the parameters you reported on in HW5 for Closed Hashing (largest cluster, farthest drift, etc.)?

4. Extension to other purpose/additional features (Suggestions)

Report on performance differences with table sizes of (1, 10, 1000, 10000) for “Moby-Dick”.

How did you implement Insert\_Key? Does it matter whether you added words to the heads or tails of lists? How could you tell?

Did you implement delete? Could you use that to report on the words in “Moby-Dick” that are not in “War and Peace” and/or vice versa?

We’ll discuss possibilities over the next week or so.

5. Written Report / Summary

Each group will document each of the steps above in a written report[[1]](#footnote--1)∇ describing/crediting the development of the group’s code, detailing the steps taken to test the code and insure its correctness, delineating the implementations strengths and weaknesses vis-à-vis the method(s) compared and contrasted to it, and depicting the intricacies of its extension.

6. Video Presentation

Each group will submit a short[[2]](#footnote-0)± video encapsulating the creative process—running code, running commentary, and running for the exits. Creativity and playfulness encouraged—nay, insisted upon!—here as well.

In this crazy term we are all picking up unexpected skillz.

The aim here is to have some fun both working out a new algorithm AND reporting on it.

1. ∇ Two pages is surely too short and eight pages likely too long, not counting considerable possible output which may be included as an appendix. [↑](#footnote-ref--1)
2. ±± Three minutes is surely too short and eleven minutes likely too long. [↑](#footnote-ref-0)