Algorithms Homework 4

Liam Dillingham

October 18, 2018

An important data structure used that may differ from other student's implementations is the Java TreeSet. Since the keys must all be unique, I wanted a way to ensure that. So I used a treeset to contain the randomly generated keys, and used a random integer generator with the constrains where any x produced by the generator falls in the range $0 \leq x \leq 100000$.

Then, within a while loop, I randomly generated keys and tried to insert them into the treeset until the size of the treeset equaled 950. Then I partitioned the set of keys into a group of 900, and of 50, as per the assignment.

I think it is valuable to note that we used the same auxiliary hash function for all necessary hashing algorithms. As stated in 11.4-1, the hashing function is: $h'(k) = k$, or the identity function.

For the linear-probing function, the equation looks as follows: $h(k, i) = (h'(k) + i) \mod m$ where i is the current number of collisions, and m is the table size.

For quadratic-probing, we use the function: $h(k, i) = (h'(k) + c_1 * i + c_2 * i^2) \mod m$, where $c_1 = 1$, and $c_2 = 3$ as per problem 11.4-1.

For double-hashing, we have the equation: $h(k, i) = (h_1(k) + i * h_2(k)) \mod m$, where $h_1(k) = k$, and $h_2(k) = 1 + (k \mod (m - 1))$.

Now for the testing results. For each hashing function, we reuse the same data for both initial hashing, and testing. However, we use a different/new table for each function.

For the analysis, we use the paritioned dataset of 50 elements, and measure the total number of probes (number of times the function tries to insert a key into the table), and the number of collisions (number of times the function tries to insert and fails). It should be noted that the number of collisions will always be 50 fewer than the number of probes. Below are the results:

```
# Probes for LINEAR hashing (analysis data): 2077
# Collisions for LINEAR hashing (analysis data): 2027

# Probes for QUADRATIC hashing (analysis data): 592
# Collisions for QUADRATIC hashing (analysis data): 542

# Probes for DOUBLE hashing (analysis data): 520
# Collisions for DOUBLE hashing (analysis data): 470
```