CS 3410 - Homework 05

Due date: see course Schedule and Blackboard.

Overview

You will conduct an experiment to determine the average number of cells examined during a random add to a HashSet using Linear Probing (B-version) or Separate Chaining Hashing (A-version) with a load factor of 0.4995. You will compare your results to theoretical values.

Requirements – B Version (85 points maximum)

1. Obtain the download, *hw5.zip*.
2. (40 points) Implement linear probing. The current HashSet implements quadratic probing. The change to linear probing is extremely easy.

   Hint: All probing is in the *findPos* method in HashSet.java.

3. Add code to count the number of cells examined during an *add* method call.

   Hint: the simplest thing to do is put in a public global counter variable, say *Probe_Count*. Then, set its value to 1 in *findPos* and increment it as appropriate in the correct place(s) in *findPos*.

4. (35 points) Conduct the following experiment.

    1. Repeat 100 times:
        a. Create a HashSet with an initial size of 1009 (prime).

           Hints:
            I. See the *HashSet* constructor to determine the variable that is controlling the initial table size.
            II. See *SetDemo.java* for an example of creating a hash set and adding a value.

        b. Add 503 random integers between 0 and 10,000. This is a load factor of 49.85%. (Here, you don't care about the counts of cells examined). A HashSet will not allow duplicates, so make sure you do indeed add 503 integers.

           Hint: Check the return type of the *add* method to make sure an add is successful

        c. Add another random value and record the number of probes (for this single add), *e.g.* Probe_Count. Again, make sure the value is added before recording the number of probes.

           Hint: Probe_Count should usually be a value between 1 and 5 for a single add.

        d. Increment the total number of probes: pTotal += Probe_Count

    2. Compute the average number of probes: pAverage = pTotal / 100

5. (10 points) Compare your result to two theoretical values: (a) no clustering, (b) clustering.

   Hint: there are formulas for these in the text (and notes).

Requirements – A Version (120 points maximum)

1. Read B version to get some of the hints. They will be useful for the A version.
2. Obtain the download, *hw5.zip*.
3. (65 points) Implement separate chaining hashing. Think carefully about this before you code.
4. Add code to count the number of cells examined during an *add*.
5. (35 points) Conduct the following experiment.

    1. Repeat 100 times:
        a. Create a HashSet with an initial size of 1009 (prime).
        b. Add 503 random integers between 0 and 10,000. This is a load factor of 49.95%. (Here, you don't care about the counts of cells examined). A HashSet will not allow duplicates, so make sure you do indeed add 503 integers.
        c. Add another random value and count the number of probes (for this single add), p. Again, make sure the value is added.
        d. pTotal += p
    2. pAverage = pTotal / 100

6. (10 points) Compare your result to two theoretical values: (a) no clustering, (b) clustering.
7. (10 points) Do the B version of the assignment.

Deliverables

1. A Word document in this format:
    a. Title page:
        CS 3410 – HW 05
        Name
        Date
    b. Answers to question 5 above.

2. All Code.