

Measurements at the end

Performance of Hashtables - Evaluation

If space is not an issue, which produces the fastest insert/search functions, and under what conditions?

If space is not an issue then we would look at the results from the load factor of 0.5 from the Load and Performance tests. From these tests it can be seen that the Sequential Chaining Hashtable is the best at inserting in words and definitions as it has the lowest amount of probes done when inserting the whole lexicon dictionary. But Quadratic is better at finding definitions as the average amount of probes taken to find a word is 0.57 probes which means half the time it just finds a position for a word instantly without probing. This is only true for the 0.5 load factor as if your list is greater than this then the Quadratic becomes exponentially more expensive. The Sequential Chaining becomes better than Quadratic at the 0.75 load factor which makes the Quadratic Probing rather risky.

Which technique generally performs best?

Generally, the Sequential Chaining hashtable performs the best from the 3 hashtables except when searching at the 0.5 load factor level where the other hashtables end up being better than Sequential Chaining. This is bad as with rehashing we can keep the load factor down so the other hashtables efficiency at searching is helpful even at the 0.5 level. But Sequential Chaining is widely better across other conditions making it the safer but still efficient option. The Sequential Chaining's leading efficiency at inserting is also very important and pushes it to being a very good option when choosing the best hashtable.

Allowing that the investigation has not considered the amounts of memory used, using estimates, which offers the best combination of speed and efficient storage?

Even though not tested, it seems the Sequential Chaining hashtable would be very expensive in memory as it has to store a linked list in every position in its table that has been filled. This is worse than the other hashtables as with the other hashtables, the word object is kept directly at an index. Only the memory storage of the word object will be kept while the Sequential Chaining has to keep the word and the link to the next word.

In conclusion, the Sequential Chaining seems to perform the best from the other tables at most load factors. Even though the Sequential Chaining is probably the most costly memory wise, it has a relatively good search functionality at any load factor and has a very efficient insert function so it can be deduced that it performs the best overall. Sequential Chaining also has the worst case of having everything inserting into it, get the same hashkey and thus have the whole dictionary just one linked list in the hash table. This is a very rare case and can easily be avoided with a good hash function though.

Description of how Measurements were obtained:

Two probe count instance variable was kept in each hashtable class. One probe count is for loading/inserting probes and the other is for search probes. The load probe counter gets increased only when a hashtable has to use its probing methodology to find a different position for a word. This happens when the calculated hash function gives a position in the table is already full so whenever a hashtable has to use a different hash value, it is probing. So we add to the probe counter everytime a probe needs to be done. The Search probes are very similar but we only add to that variable for the amount of times a hashkey needs to be changed to find a word. Including the amount of probes it takes to find out a word doesn't exist.

Measurements:**Loading/Inserting:**

Total number of probes used to insert all of lexicon words and definitions

LP: Total number of Probes is 2818 for a load factor 0.5

QP: Total number of Probes is 2194 for a load factor 0.5

SC: Total number of Probes is 952 for a load factor 0.5

Total number of probes used to insert all of lexicon words and definitions

LP: Total number of Probes is 7404 for a load factor 0.75

QP: Total number of Probes is 4560 for a load factor 0.75

SC: Total number of Probes is 1367 for a load factor 0.75

Total number of probes used to insert all of lexicon words and definitions

LP: Total number of Probes is 198753 for a load factor 1.0

QP: Total number of Probes is 27506 for a load factor 1.0

SC: Total number of Probes is 1838 for a load factor 1.0

Percentage differences

Quadratic and Linear Probing : 28% for a load factor 0.5

Sequential chaining and Quadratic Probing : 130% for a load factor 0.5

Sequential chaining and Linear Probing : 196% for a load factor 0.5

Quadratic and Linear Probing : 62% for a load factor 0.75

Sequential chaining and Quadratic Probing : 234% for a load factor 0.75

Sequential chaining and Linear Probing : 442% for a load factor 0.75

Quadratic and Linear Probing : 623% for a load factor 1.0

Sequential chaining and Quadratic Probing : 1397% for a load factor 1.0

Sequential chaining and Linear Probing : 10714% for a load factor 1.0

Searching:

The total average number of probes over 100 trials to search for a different random 100 words at each trial for load factor 0.5

Searching in LP is total: 78 Avg per word: 0.78

Searching in QP is total: 57 Avg per word : 0.57

Searching in SC is total: 110 Avg per word: 1.1

Percentage difference

Linear Probing and Quadratic Probing : 37%

Sequential Chaining and Quadratic Probing : 48%

Linear Probing and Sequential Chaining : 29%

The total average number of probes over 100 trials to search for a different random 100 words at each trial for load factor 0.75

Searching in LP is total: 273 Avg per word: 2.73

Searching in QP is total: 135 Avg per word : 1.35

Searching in SC is total: 123 Avg per word: 1.23

Percentage difference

Linear Probing and Quadratic Probing : 102%

Sequential Chaining and Quadratic Probing : 10%

Linear Probing and Sequential Chaining : 122%

The total average number of probes over 100 trials to search for a different random 100 words at each trial for load factor 1.0

Searching in LP is total: 77740 Avg per word: 777.4

Searching in QP is total: 75216 Avg per word : 752.16

Searching in SC is total: 139 Avg per word: 1.39

Percentage difference

Linear Probing and Quadratic Probing : 3%

Sequential Chaining and Quadratic Probing : 54012%

Linear Probing and Sequential Chaining : 55828%