

Data Structures

Hash Tables

Cain Susko

Queen's University
School of Computing

March 7, 2022

Recall

python or other languages have different names for hash tables. For example, in python, a hash table is a dictionary.

Also, recall the 2 most important operations of the map ADT are:

- `get()`
- `put()`

Hash tables implement a map ADT. Hash tables use direct addressing using a **hash** function and a key. The hash function is bounded within an array size using the modulus operation. for example:

$$\text{int bucket} = \text{input} \% \text{arraySize}$$

so a hash function could looklike:

```
def hash (int key):
    return key % 10000
```

Hash Tables

there are 2 types of proper hash tables. Open Hash Tables and Closed Hash Tables

Closed Hash Table A closed hash table is called such because data is stored in a fixed length array, there are a fixed number of ‘buckets’ in which you can store values. We cannot add new buckets with inserting or ‘probing’ in a closed hash table the algorithm would be as follows:

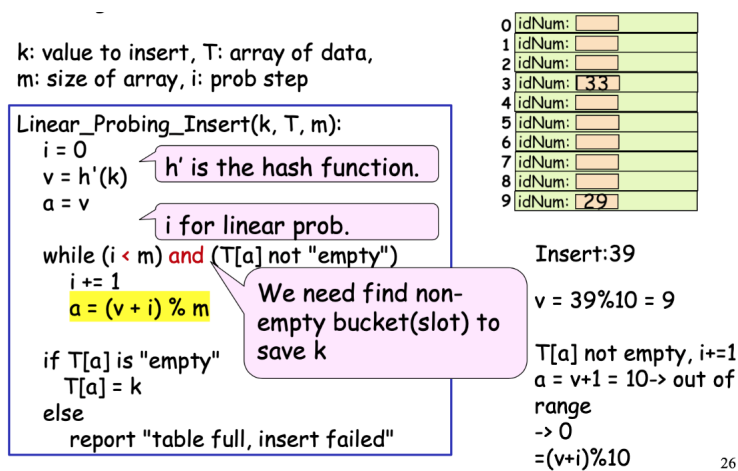
- first, we use the has function to find the correct bucket
- if the target bucket is empty, we store the value there
- however, if the bucket is full, scan down the array from the bucket untill you find a empty one

the Algorithm for searching is also quite similar:

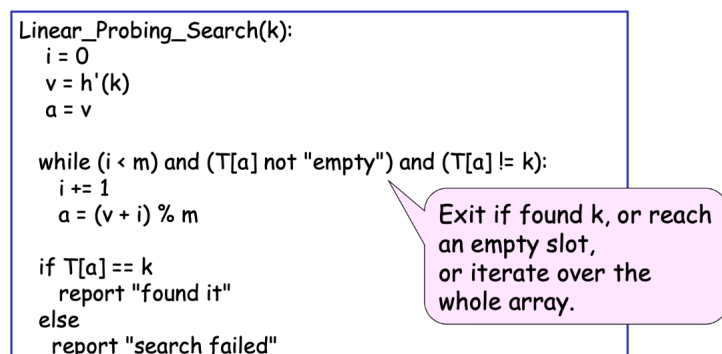
- We compute the target bucket using the hash function

- we check that bucket for our value. if we find it, return true
- if we don't find the value, we 'probe' linearly down the array until we either find our value or hit an empty bucket
- if while probing, we reach an empty bucket (the end) without finding the value, it means the value is not in the hash table

the insert function (using linear probing) would look something like this:



And the Search function would operate like so:



So, how do we delete a item from the Hash Table?

For example, lets delete 65 from the below hash table. To delete it, we can use a 'delete flag' to show a element is deleted. this may cause the insert and within the table to take longer as you delete more and more and more of the table is blocked out by the delete flag.

Open Hash Tables A open hash table does not have a fixed number of buckets, but this means that the hash function does not always create numbers within the bounds of the open hash table. the way it works is instead of storing the values directly in the array, each array bucket points to a linked list of values.

For example, to insert a new item:

- compute the bucket index using the hash function (like always)
- add the item to the linked list at *array[bucket_array]*

Assignment 3

it is a web search engine

there are 2 parts of the engine:

- Docs: 1 to n
- Quereys: 1 to n

we have to majke a WebPageIndex class that maps the lines of docs to an AVL tree

we also have to make a priority queue for storing the Queries and words.

This shit is going to be painful cause ive got no idea how to do any of that.