CSCI 2270 Lecture Notes

3/4/19

Introduction to Hash Tables

* Hash Tables is a data structure that uses mapping to assign a “record” to a unique index of an array.
* given a key, a hash function generates a unique index
* Two Components
  + an array for storing the records
  + hash function for generating the unique code given a key
    - Hash function is repeatable
      * hash(“Billy) = 7
      * will ALWAYS be a 7
    - unique code is used as an array index

Hash Functions

* one of the simplest is to sum the values in the string key, then mod(%) by array length (table size)

hashSum(key, keyLength, tableSize)🡨 NOT a great hash function. Will probably result in collisions

sum = 0

for i=0 to keyLength – 1

sum = sum + key[i]

return sum%tableSize

Store Records in a hash table

* Calculate index value (hash the key)
* Write the data to hash table at the index we calculated with hash

Retrieve Records

* use search key to calculate index with hash function
* read from the index

What is the cost of store? = O(1)

retrieve? = O(1)

Collision

* hashSum(“Go Cat Go.”) = 754
* hashSum(“Go Dog, Go”) = 754
  + both of these result in the same index, but each array location can only store a single record