**COMP1201 Assignment 2**

Q1:

*What we are given:*

* Hash function with separate chaining.
* Takes an integer **(n < 100 000)**.
* Calculates the value as **(2d1 + 3d2 + 5d3 + 7d4 + 11d5) % 47**, where d1 is the most significant digit and d5 is the least significant digit.
* We enter **2000** numbers.

*What we need to prove:*

* Prove that there exists a number **x** **(0 <= x <= 46)** for which we can find **at least 43** numbers among the **given** **2000**, whose **hash** **value** is exactly **x**.

*Proof:*

First, when a hash table has separate chaining, it mean that we build a singly-linked list at each table entry. If the keys for two elements are the same, then both of the elements are added to the corresponding singly-linked list.

The given problem can be easily coded using Java:

//Create the new HashMap, which will be implemented using separate chaining.  
HashMap<Integer, ArrayList<Integer>> myMap = new HashMap<>();  
Random rand = new Random();  
  
int key = 0;  
int value = 0;  
  
//Add 2000 numbers.  
for (int i = 0; i < 2000; i++) {  
 //Generate the 2000 random numbers.  
 value = rand.nextInt(100000);  
  
 //Isolate the different digits, in order to calculate the hashing value.  
 int firstDigit = (value / 10000);  
 int secondDigit = (value / 1000) % 10;  
 int thirdDigit = (value / 100) % 10;  
 int fourthDigit = (value / 10) % 10;  
 int fifthDigit = (value % 10);  
  
 //Add the key and value to the map.  
 key = (2 \* firstDigit +  
 3 \* secondDigit +  
 5 \* thirdDigit +  
 7 \* fourthDigit +  
 11 \* fifthDigit) % 47;  
  
 //Lambda expression to add the integer to its specified ArrayList.  
 myMap.computeIfAbsent(key, k -> new ArrayList<>()).add(value);  
}  
  
//Iterate through the map to see the values.  
int i = 1;  
Integer lastKey = 0;  
ArrayList<Integer> numberOfValues= new ArrayList<>();  
  
for (java.util.Map.Entry<Integer, ArrayList<Integer>> entry : myMap.entrySet()) {  
 Integer myKey = entry.getKey();  
 ArrayList<Integer> myValues = entry.getValue();  
 for (Integer singleValue : myValues) {  
 //Print out everything.  
 //The counter is there so we can easily see the number of elements under each key.  
 if(lastKey == myKey) {  
 //System.out.println("key : " + myKey + " value : " + singleValue + " number : " + i);  
 i++;  
 }  
 else {  
 i = 1;  
 //System.out.println("key : " + myKey + " value : " + singleValue + " number : " + i);  
 lastKey = myKey;  
 i++;  
 }  
 }  
 //System.out.println("Key: " + myKey + " <-> " + "Number of values: " + i);  
 numberOfValues.add(i);  
}  
  
//Display the amount of values under each key, sorted, so we can prove that  
//that there exists a number x between 0 and 46 for which we can find at least 43  
//numbers among the given 2000, whose hash value is exactly x.  
Collections.*sort*(numberOfValues, Collections.*reverseOrder*());  
for(Integer number : numberOfValues) {  
 System.*out*.println(number);  
}

If we run this code, we can see that there is at least one number, **x**, between 0 and 46, for which there are at least 43 numbers, from the 2000, whose hash values is **x**.

Now, if we try to prove this mathematically, we can use the *pigeonhole principle*.

First, given the *hash code*, and the given *maximum value* we can have, the *values* of the *keys* may vary from 0 (included) to 46 (included). Since we input 2000 numbers (pigeons), and we have 47 key slots (pigeonholes), we see that there will be at least one key (pigeonhole) with at a minimum of 43 values (pigeons) under it. We can see that by dividing the pigeons by the pigeonholes (2000 / 47), and we get 42,5531914893617‬. When we the ceiling function (as per pigeonhole principle), we can see that (**⌈**42,5531914893617**⌉** = 43)there is a pigeonhole with at least 43 pigeons in it.

Q2:

*What we need to do:*

* Design a data type, that supports the following operations:
  + Insert an element (**O (log (n))** time complexity).
  + Delete the biggest element (**O (log (n))** time complexity).
  + Delete the smallest element (**O (log (n))** time complexity).
  + Find the biggest element (**O (1))** time complexity).
  + Find the smallest element (**O (1))** time complexity).
* Design the data type using two heaps.

*Creating the data type:*

Before we start adding elements to any data types, we create two heaps. One is a **min heap** (smallest element is the root of the heap), and the other is a **max heap** (biggest element is the root of the heap). Because we want to keep the time complexity of deleting O (log (n)), we need to create another data structure to hold all our elements, so we can easily delete an element from both our heaps simultaneously. So we create a **doubly linked list**, which will contain all input items and indexes of the corresponding min and max heap nodes. We will use a doubly linked list as its insertion and deletion time complexity are O (1).

Representation of idea:

Insert element “**1**”.

* First we encode the inserted value. For the sake of simplicity, we just multiply the value by 100. (So the value that will be stored in the heaps will be **100**)

Max Heap

Min Heap

0

0

Doubly Linked List

0 1 0

Min Heap Indexes

Max Heap Indexes

*Next Insert*

Insert element “**2**”.

* Again we encode it.

0

0

1

1

1 2 0

0 1 1

*Next Insert*

Insert element “**3**”.

2 1 0

1 2 2

0 1 1

2

1

0

2

1

0

How it works:

**Inserting an element:** We add the element to the start of the doubly linked list **(O (1))**. Then we insert the element to the **min** and **max** **heaps** **(O log (n))**. Thus making the overall complexity be **O (log (n))**.

Pseudocode for inserting an element:

//Inserts an element into both heaps and the linked list

void insert(int key) {  
 Node node = new Node();  
 node.data = key;  
 node.minHeapIndex = minHeap.insert(node);  
 node.maxHeapIndex = maxHeap.insert(node);  
 list.insertAtBeginning(node);  
}

//Perculates the element through the heap.

//Helper method  
int fixUpwards(int i) {  
 while (i != 0 && arr[i].data > arr[parent(i)].data) {  
 swap(arr, i, parent(i));  
 i = parent(i);  
 }  
  
 return i;  
}

//Inserting an element into the min heap.  
int insert(Node node) {  
 if(size == capacity) {  
 System.*out*.println("Heap overflow");  
 return -1;  
 }  
  
 arr[size] = node;  
 arr[size].minHeapIndex = size;  
 size++;  
 return fixUpwards(size-1);  
}

//Inserting an element into the max heap.  
int insert(Node node) {  
 if (size == capacity) {  
 System.*out*.println("Heap overflow");  
 return -1;  
 }  
  
 arr[size] = node;  
 arr[size].maxHeapIndex = size;  
 size++;  
 return fixUpwards(size - 1);  
}

//Inserts the element at the begging of the linked list.  
void insertAtBeginning(Node node) {  
 node.prev = null;  
 node.next = head;  
  
 if (head != null)  
 head.prev = node;  
  
 head = node;  
}  
int findMax() {  
 return maxHeap.getMax().data;  
}

**Finding the smallest element:** We look at the root of the min heap. This is a O (1) operation.

//Finds the smallest element.  
int findMin() {  
 return minHeap.getMin().data;  
}

**Deleting the biggest element:** We get the biggest element using the findMax method. Then we get its address to find the node in the doubly linked list. From there we get the position of the element in the max heap. We remove the element from the three places (min heap, max heap and the linked list). Deleting the element from the doubly linked list has a O (1) time complexity. Removing the element from the heaps has a time complexity of O (log (n)). Which means that we achieve the needed overall time complexity.

//Deletes the biggest element.  
int deleteMax() {  
 Node node = maxHeap.getMax();  
  
 int result = node.data;  
  
 int minIndex = node.minHeapIndex;  
 list.deleteNode(node);

//extractMax() is a helper method that returns the root of the max heap.  
 maxHeap.extractMax();  
 minHeap.delete(minIndex);  
  
 return result;  
}

**Deleting the smallest element:** We get the smallest element using the findMin method. Then we get its address to find the node in the doubly linked list. From there we get the position of the element in the min heap. We remove the element from the three places (min heap, max heap and the linked list). Deleting the element from the doubly linked list has a O (1) time complexity. Removing the element from the heaps has a time complexity of O (log (n)). Which means that we achieve the needed overall time complexity.

//Deletes the smallest element  
int deleteMin() {  
 Node node = minHeap.getMin();  
  
 int result = node.data;  
  
 int maxIndex = node.maxHeapIndex;

//extractMin() is a helper method that returns the root of the min heap.  
 minHeap.extractMin();  
 maxHeap.delete(maxIndex);  
 list.deleteNode(node);  
  
 return result;  
}