Data Structures and Algorithms Solutions to Quiz 5

1. A hash function
a) Must map records with unique keys to unique slots in the hash table
b) Must use all the hash table slots randomly
c) Must make potential use of all the hash table slots
d) All of the above
e) None of the above
a) is wrong because sometimes records with unique (different) keys map to the same slot. b) is also wrong because the hash table slots are not visited randomly. They are visited according to some mathematical formula (hash and probe functions). **c)** is the correct answer because every slot in the table should be a candidate for insertions and searching. This means that sooner or later, a given slot will be visited.

2. Given a Hash table of size `K` and `N` records to be placed in the table (`N` ≤ `K`), the following is a valid hash function (works but may not be optimal):
a) `int h(int x) { return x;}`
b) `int h(int x) { return x%N;}`
c) `int h(int x) { return x%K;}`
d) a) or b)
e) a) or c)
f) b) or c)
g) All of the above
f) is the correct answer because in both cases (b) and c)), the hash function returns an index that is less than the size of the table. In the case of b), this is not a "good" hash function since some of the slots will never be visited – however, it works.

3. Given the following Hash function `int h(int x) {return y=f(x);}`, then
a) `x` is unique for different records
b) `y` is unique when `x` is unique
c) a) and b)
d) None of the above
b) is wrong because unique keys could hash to non-unique slots. However, the keys themselves are unique. Hence, answer **a)** is correct.

4. Using a Hash function `h(k)` to search for a given record, the search is stopped when
a) A full slot is encountered and the key of the record matches the search key.
b) A Tombstone is encountered
c) An empty slot is encountered
d) All of the above
e) b) and c)
f) a) and c)
g) a) and b)
Answer **f)** is correct. A Tombstone does not stop the search.

5. Hashing inserts a record in the slot that
a) is empty
b) has a Tombstone
c) has a record with a matching key
d) All of the above
e) b) and c)
f) a) and c)
g) a) and b)
Answer **g)** is correct. A Tombstone indicates that the slot is available for insertion.

6. Hashing is most appropriate for:
a) Range queries
b) Exact-match queries
c) Less than/Greater than value queries
d) All of the above
Hashing is most-appropriate for Exact-match queries: Answer **b)** is the correct one

7. The following Probe function `int p(int x, int i) {return 4*i;}` works
a) for Closed Hashing
b) for Open Hashing
c) for both Open and Closed Hashing
d) for Neither, Open Hashing nor Closed Hashing
The above function does not work for bucket hashing nor open hashing. answer **a)** is the correct one.

8. If `k` is the record's key, `M` is the size of the Hash table, and `i` is the probe sequence index, the following is a valid C++ expression used to determine a Hash Table slot
a) `pos=(h(k)+p(k,i))%M`
b) `pos=h(k)%M+p(k,i)`
c) `pos=h(k)+p(k,i)%M`
d) `pos=h(k)%M+p(k,i)%M`
Answer **a)** is the ONLY correct answer. The modulo operator must be used for the sum of both the hash function and the probe function in order to ensure that the resultant slot index does not exceed the boundary of the hash table.

9. Secondary Clustering occurs when
a) records tend to be inserted in certain areas in the Hash table more than in others
b) keys hash to same slot, then they follow the same probe sequence
c) keys hash to same slot, then they follow probe sequences that have common slots
d) keys hash to different slots, then they follow the same probe sequence later
Secondary clustering is defined exactly as what answer **b)** states.

10. Secondary Clustering is associated with
a) Linear Probing
b) Pseudo-Random Probing
c) Quadratic Probing
d) All of the above
Secondary Clustering can occur for all above-three probing techniques. **d)** is correct

11. With Pseudo-random probing,
a) Same probe sequence is added to Hash function outputs for different records
b) Different probe sequences are added to Hash function outputs for different records
c) Insertion and Searching use the same sequence
d) Insertion and Searching use different sequences
e) a) and c)
f) a) and d)
g) b) and c)
h) b) and d)
Answer **e)** is the correct answer because for Pseudo-random probing, a) and c) apply.

12. Suppose we use quadratic probing to search for records. We have two records whereby the output of the hash function for record 1 is 15 and the output of the hash function for record 2 is 19. Suppose we have a Hash table of size 50 and it takes 8 searches for each record to be found, then the probe sequences are
a) `[15,16,19,24,31,40,01,14]` and `[19,20,23,28,35,44,05,18]`
b) `[16,19,24,31,40,01,14,29]` and `[20,23,28,35,44,05,18,33]`
c) `[15,16,20,29,45,20,06,05]` and `[19,20,24,33,49,24,10,09]`
d) `[15,16,19,24,31,40,51,64]` and `[19,20,23,28,35,44,55,68]`
e) `[15,16,20,29,45,70,106,155]` and `[19,20,24,33,49,74,110,159]`
With quadratic probing, the probe sequence is 0, 1, 4, 9, 16, 25, 36, … This sequence is added to the home positions for both records and then its modulo(50) is taken. Hence, answer **a)** is the correct one.

13. Suppose we use double hashing to generate probe sequences for a Hash table with 100 slots. If $h1(k)$ is the first hash function and $h2(k)$ is the second hash function, then when have 3 records to be searched for such that $h1(k1)=337$, $h2(k1)=7$, $h1(k2)=145$, $h2(k2)=27$, $h1(k3)=97$, $h2(k3)=43$, then the probe sequences for $k1$, $k2$, and $k3$ are
a) `[37,44,51,58,65],` `[45,72,99,26,53],` and `[97,40,83,26,69]`
b) `[37,44,58,79,7],` `[45,72,26,7,15],` and `[9,7,40,26,55,27]`
c) `[37,44,51,58,65],` `[45,72,99,126,153],` and `[97,140,183,226,269]`
d) `[37,44,51,58,65],` `[45,72,99,128,53],` and `[97,140,183,126,169]`
With double-hashing, the slot index is pos=modulo($h1(k)+i* h2(k)$), 100), i=0,1,2,3,…
If you do the math, then you'll find that answer **a)** is the correct one.

14. Suppose we use bucket Hashing with M=5 (M is the number of home positions), the number of slots per bucket is 5 (including the home slot), and the size of the overflow bucket is 50. Suppose we have the following hash function $h(k)=k\%M$. if we have the following keys:

`37,44,58,79,17,45,72,26,7,15,77,40,36,55,27,97,57,66,54,9`

Then the number of records in the overflow bucket will be

a) 0
b) 1
c) 2
d) 3
e) 4
f) 5

For the above keys, we compute the home positions (pos=modulo(k,M)) and get the corresponding slots: 2, 4, 3, 4, 2, 0, 2, 1, 2, 0, 2, 0, 1, 0, 2, 2, 2, 1, 4, 4

As we can see, 8 records hash to home position 2. However, for each home position is part of a bucket that has 5 slots. Hence, 3 records will have to go into the overflow bucket. Therefore, answer **d)** is the correct one.

15. Suppose we use closed Hashing with M=50. We have the hash function $h(k)=11+k$ and the probe function $p(k,i)=i$. . If we insert the following keys into an empty table

`37,44,58,79,17,45,72,26,7,15,77,40,36,55,27,97,9,57,66,54`

and later search for the 20 records that have these keys. The maximum number of table slots that need to be searched before finding any of these records is

a) 1
b) 2
c) 3
d) 4
e) 5

For the above keys, the corresponding home positions are (h=11+k):
48, 55, 69, 90, 28, 56, 83, 37, 18, 26, 88, 51, 47, 66, 38, 108, 20, 68, 77, 65

Then when we compute the slot position (pos=modulo(h+i, M), i=0,1,2,…). That is, we start with i=0 and if we detect that the slot is filled, we increment i until we find an available slot in the range 0 to 49:
48, 5, 19, 40, 28, 6, 33, 37, 18, 26, 38, 1, 47, 16, 39^{1}, 8, 20, 21^{3}, 27, 15

In the above, one record hashed to slot 38, which was full, so slot 39 was used since it was available(in red). One record hashed to slot 18, which was full, then 19, was full, 20 was full, and then 21 was available (in pink). Therefore, when searching for the record whose key is 68, it will take 4 searches to find it. As a result, answer **d)** is the one.

16. Primary Clustering is associated with

a) Linear Probing
b) Pseudo-Random Probing
c) Quadratic Probing
d) All of the above

Primary Clustering is associated with Linear Probing. Answer **a)** is the correct one.