Data Structures and Algorithms Assignment 6 Max Marks 40

Question 1 (18 marks)

Suppose you have a hash table of size 19, the keys are words, and the hash map is defined as follows: Each letter is assigned a number according to its position in the alphabet, ie

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	P	q	r	s	t	u	v	w	x	y
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

and the primary hash function is "x modulo 19", where x is the number corresponding to the first letter of the word. Why is this hash function not ideal?

Suppose instead you have a hash table of size 13 and the primary hash function is "x modulo 13", where x is the sum of the numbers corresponding to all the letters in the key word. Insert the following list of words into an initially empty hash table using linear probing:

[computer, science, in, birmingham, dates, back, to, the, sixties]

What is the load factor of the resulting table, and how many collisions occurred?

What is the effort (i.e. number of comparisons) involved in checking whether each of the following words are in the hash table: teaching, research, admin?

Show what the resulting hash table would look like if direct chaining had been used rather than linear probing. Now what is the effort (i.e. number of comparisons, not including the processing of NULL pointers) involved in checking whether each of the following words are in the hash table: teaching, research, admin?

Question 2 (12 marks)

A call centre uses a hash table with open addressing to store customer orders taken during its working hours of 7am to 11pm every day of the year. Orders are kept in the hash table for 5 years and then deleted. It currently has about 5 million entries, and handles about 1 million transactions per year (i.e. between 2 and 3 a minute).

What size hash table would you suggest?

The performance of the hash table has degraded. Why might this have happened?

Suppose someone suggested copying the entries in the hash table into a new hash table, and it takes 2 milliseconds to copy one item. Would that be a sensible thing to do?

Question 3 (5 marks)

Consider a hash table of size seven, with starting index zero, and a hash function (3x + 4)mod7. Assuming the hash table is initially empty, which of the following is the contents of the table when the sequence 1, 3, 8, 10 is inserted into the table using closed hashing?

Question 4 (5 marks)

Write an efficient function to return maximum occurring character in the input string e.g., if input string is "test" then function should return 't'.

note: use hashing