2.6 Hash tables

Name: _____

Class: _____

Date: _____

Time: **22 minutes**

Marks: **14 marks**

Comments:

Q1.

A dictionary is an abstract data type that allows pairs of values to be associated with each other. For example, an English-French dictionary might associate English words with their translations into French. In such a dictionary, "Apple" would be associated with "Pomme" because "Pomme" is the French word for "Apple".

At a lower level of abstraction, a dictionary could be implemented as a data structure using a number of different methods. Two possible implementation methods are:

• **Implementation One**: As an unordered list in an array.
• **Implementation Two**: Using hashing, with the English word being passed through a hash function to calculate the position of the correct French translation in an array.

In each implementation, a record containing the English word and the equivalent French word are stored at each index in the array that is in use.

The figure below shows how an English-French dictionary containing five words could be implemented using these two methods.

Implementation One – Unordered List			Implementation Two – Hashing		
Index	English Word	French Word	Index	English Word	French Word
[1]	Apple	Pomme	[1]	Pear	Poire
[2]	Lemon	Citron	[2]	Lemon	Citron
[3]	Strawberry	Fraise	[3]		
[4]	Grapefruit	Pamplemousse	[4]		
[5]	Pear	Poire	[5]		
[6]			[6]		
[7]			[7]	Apple	Pomme
[8]			[8]	Strawberry	Fraise
[9]			[9]		
[10]			[10]	Grapefruit	Pamplemousse
New words are inserted into the array in the first available slot. The next word would be stored at position 6.			The position to store a word is calculated from the English word using a hashing function.		

(a)   Explain why, when the French translation of an English word needs to be looked up, **Implementation Two** is more time efficient than **Implementation One**.

_____

_____

_____

_____

_____

(2)

(b)   In **Implementation Two**, it is possible that the hash function could compute the same value for two different English words.

Explain what the effect of this would be, and how it could be dealt with.

_____

_____

_____

_____

_____

_____

(2)

(c)   In **Implementation Two**, both the English and French words are stored at each index in the Array. In this implementation, explain why it would not be possible to perform reliable English to French translation if only the French words were stored.

_____

_____

_____

_____

_____

_____

(1)

(Total 5 marks)

Q2.

A computer program is being developed for a car hire company. The program must store, in a file, details of the 600 vehicles that the company owns.

The records in the file will be stored and retrieved using hashing.

An alternative method that could be used instead of hashing would be to store the records in order of registration number, and use a search algorithm such as binary search for retrieval.

(a)   (i)   State **one** advantage of organising the data using hashing instead of organising the data in order by registration number.

_____

_____

_____

(1)

(ii)   State **one** advantage of organising the data in order by registration number
        instead of organising the data using hashing.

_____

_____

_____

(1)

Each vehicle is uniquely identified by its registration number. A registration number
consists of:

•   two alphabetic characters
•   followed by two numeric digits
•   followed by three further alphabetic characters.

An example registration number is DA18CFE.

The programmer has chosen the hash function below to calculate a hash value from a
registration number.

Hash value = (position in alphabet of letter at position 1 +
                position in alphabet of letter at position 2 * 10 +
                numeric digit at position 3 * 100 +
                numeric digit at position 4 * 500) MOD 1000

For the example DA18CFE the hash value would be calculated as follows:

Hash value = (position in alphabet of 'D' (4) +
                position in alphabet of 'A' (1) * 10 +
                1 * 100 +
                8 * 500) MOD 1000
            = 4114 MOD 1000
            = 114

(b)   Calculate the hash values for the following **two** registration numbers. You may use
       the space provided for working, if required.

AE21KWB

Working _____

_____

Hash value _____

KD70DAF

Working _____

_____

Hash value _____

(1)

(c)   Calculating the hash values for the two registration numbers in part **(b)** has produced a collision.

In the context of storing data in files using hashing, explain the effect of this collision and how this might be dealt with.

_____

_____

_____

_____

_____

(2)

(Total 5 marks)

Q3.

When a member of staff logs onto a computer in a particular organisation, they enter their User_name and password. The password has to be at least 8 characters long and must include both letters and numbers.

(a)   Why do organisations set rules for acceptable user passwords?

_____

(1)

The file **Password** is used by the computer system to authenticate (check) the identity and password typed at a keyboard when a user logs onto the system. The password is input to the logon program as an alphanumeric string and converted to a two-byte integer using a hashing function or algorithm before being sent across a network for authentication (checking).

(b)   Outline **three** major steps that a typical hashing function / algorithm would use to convert an alphanumeric string into a two-byte integer.

1. _____

_____

2. _____

_____

3. _____

_____

(3)

(Total 4 marks)