AQA
Realising potential

LANGLEY PARK SCHOOL FOR BOYS

MORES ET STUDIA

GCSE COMPUTER SCIENCE

GCSE (8520)

UNIT 3.8

3.8 Data compression

Content	Additional information
Explain what data compression is. Understand why data may be compressed and that there are different ways to compress data.	Students should understand that it is common for data to be compressed and should be able to explain why it may be necessary or desirable to compress data.
Explain how data can be compressed using Huffman coding. Be able to interpret/create Huffman trees.	Students should be familiar with the process of using a tree to represent the Huffman code. Students should be able to create a new Huffman tree or use a given Huffman tree to: • determine the code used for a particular node within the tree (encoding) • determine the node within a tree given its code (decoding).
Be able to calculate the number of bits required to store a piece of data compressed using Huffman coding. Be able to calculate the number of bits required to store a piece of uncompressed data in ASCII.	Students should be familiar with carrying out calculations to determine the number of bits saved by compressing a piece of data using Huffman coding.
Explain how data can be compressed using run length encoding (RLE).	Students should be familiar with the process of using frequency/data pairs to reduce the amount of data stored.
Represent data in RLE frequency/data pairs.	Students could be given a bitmap representation and they would be expected to show the frequency and value pairs for each row, eg 0000011100000011 would become 5 0 3 1 6 0 2 1.

SPECIMEN MATERIAL 2015

0 5 . 3 Sound files are stored as bit patterns. Bit patterns are often compressed.

Compress the following bit pattern using run length encoding.

```
0000 0011 1111 1000 0000 0000 0111 1111
```

[4 marks]

0 5 . 4 Shade **one** lozenge which shows the **true** statement about run length encoding:

[1 mark]

A It will always make a file smaller. ⬭

B It is most effective on data that appears random. ⬭

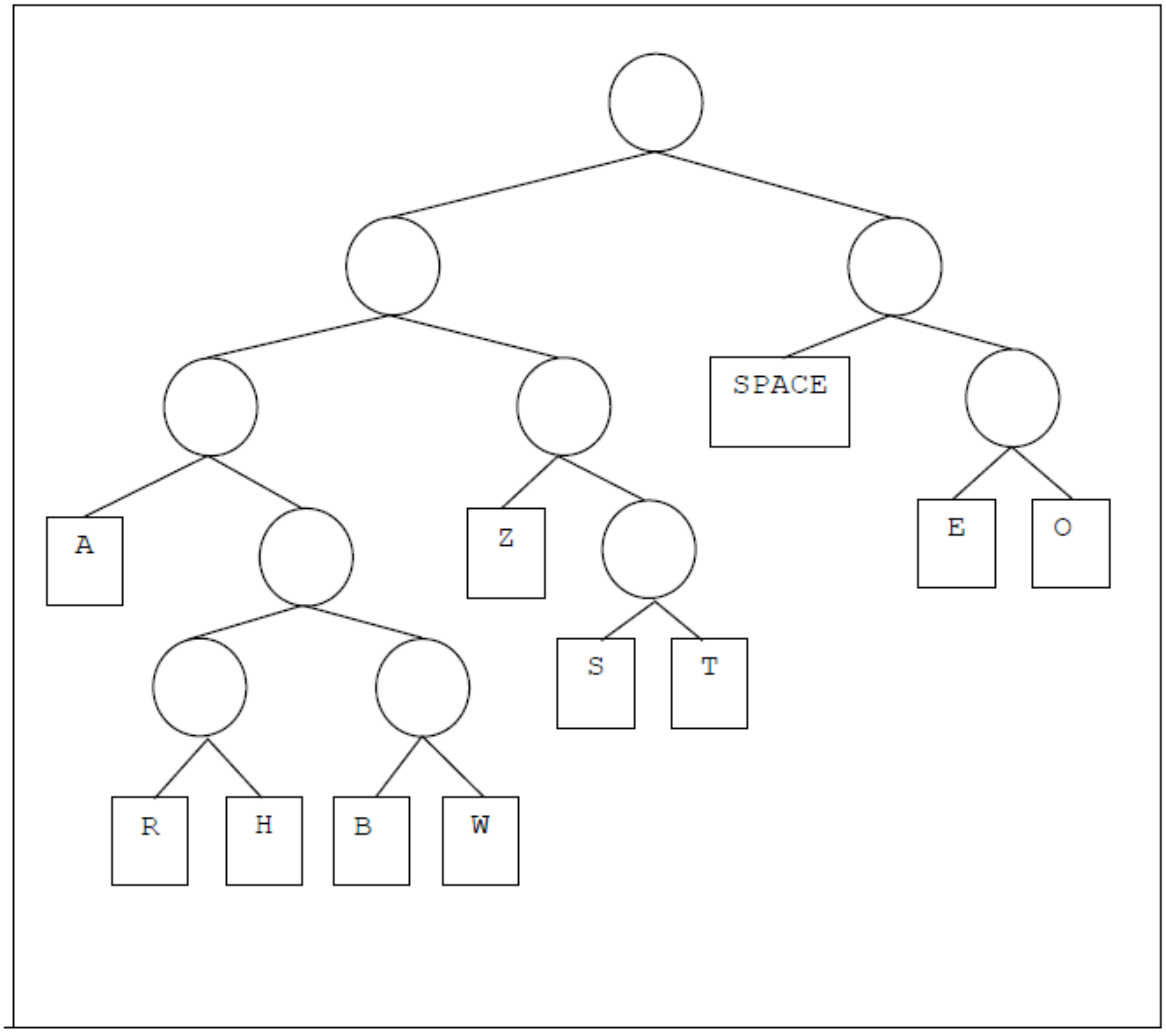
C It will not lose any of the original data. ⬭

SPECIMEN MATERIAL 2015

When data is stored in a computer it is often compressed. One method that can be used to compress text data is Huffman coding. To produce a Huffman code each character in a piece of text is placed in a tree, with its position in the tree determined by how often the character was used in the piece of text.

A Huffman tree for the text `ZOE SAW A ZEBRA AT THE ZOO` is shown in **Figure 3**.

Figure 3



Using this Huffman tree the Huffman coding for the character E would be the bit pattern 110 because from the top of the tree E is to the right, then right again and then left.

The character Z is represented by the bit pattern 010 because from the top of the tree Z is to the left, then right and then left.

0 1 . 7 Using the Huffman code in **Figure 3**, complete the table to show the Huffman coding for the characters O, SPACE and B.

[3 marks]

Character	Huffman coding
O	
SPACE	
B	

0 1 . 8 Using Huffman coding the text ZOE SAW A ZEBRA AT THE ZOO can be stored in 83 bits.

Calculate how many additional bits are needed to store the same piece of text using ASCII. Show your working.

[3 marks]

SPECIMEN PAPER 1 SUPPLEMENTAL MATERIAL 2015

1 (d) The ASCII character set uses seven bits to encode every character.

What is the total number of characters that can be encoded in ASCII?

[1 mark]

.....

.....

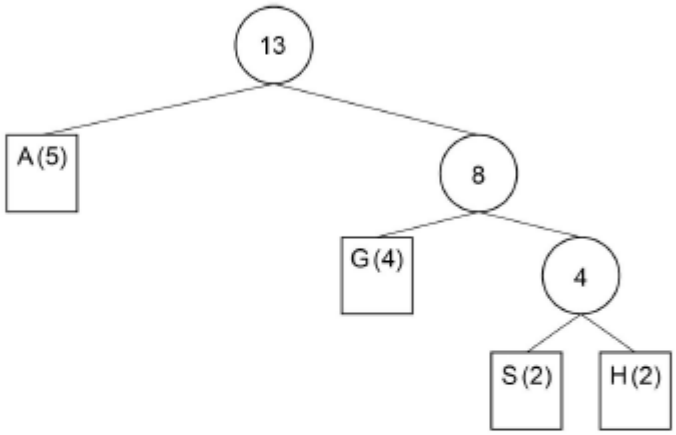
PAPER 2 JUNE 2018

1 3 The Huffman tree shown in **Figure 6** was created to encode the string shown in **Figure 5**. The frequency of each character is shown in brackets. For example, the letter A appears five times within the string shown in **Figure 5**.

Figure 5

AAGHHGGSAAASG

Figure 6



1 3 . 1 Complete the code table below for characters G, S and H for the Huffman tree shown in **Figure 6**. The code for character A has already been completed.

[3 marks]

Character	Binary code
A	0
G	
S	
H	

1 3 . 2 The string shown in **Figure 5** could also be encoded using ASCII. ASCII uses 7 bits to represent each character.

How many bits are **saved** by using Huffman coding rather than ASCII to represent the string shown in **Figure 5**?

You **must** show your working.

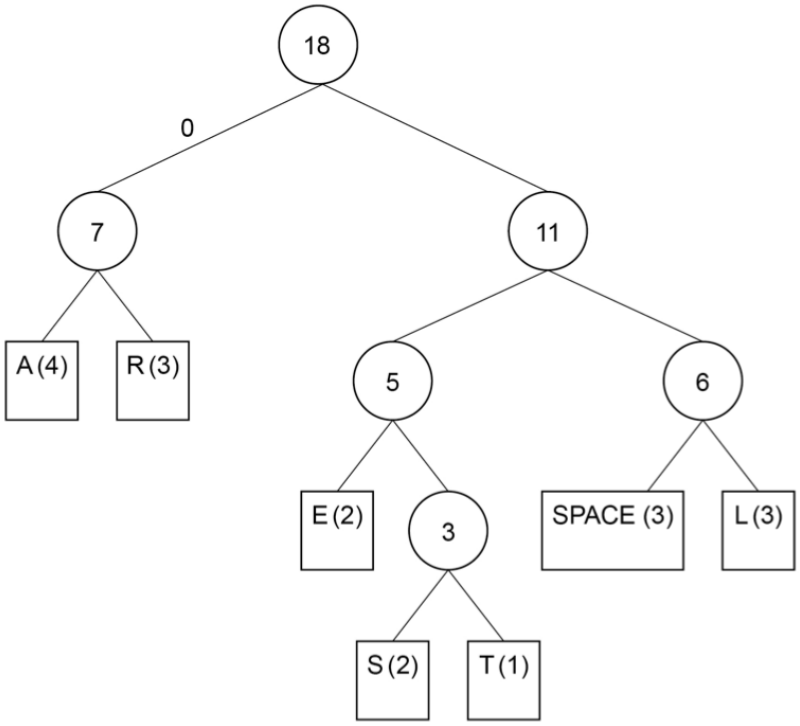
[4 marks]

Answer: _____

2019

0 8 The Huffman tree in **Figure 1** was generated for the string `ARE ALL STARS REAL`

Figure 1



0 8 . 1 Part of the string `ARE ALL STARS REAL` was incorrectly encoded as in **Figure 2** below.

Figure 2

`111100001010101 1`

What string does this encoding represent?

[1 mark]

0 8 . 2 What would be the correct binary encoding for the substring `STAR`?

Write the correct encoding below the letters in the table.

[2 marks]

S	T	A	R