

# Starter – Homework Review

**Use the answers in your notebooks to check your work. Give yourself a score and correct any mistakes using a different colour pen/font**

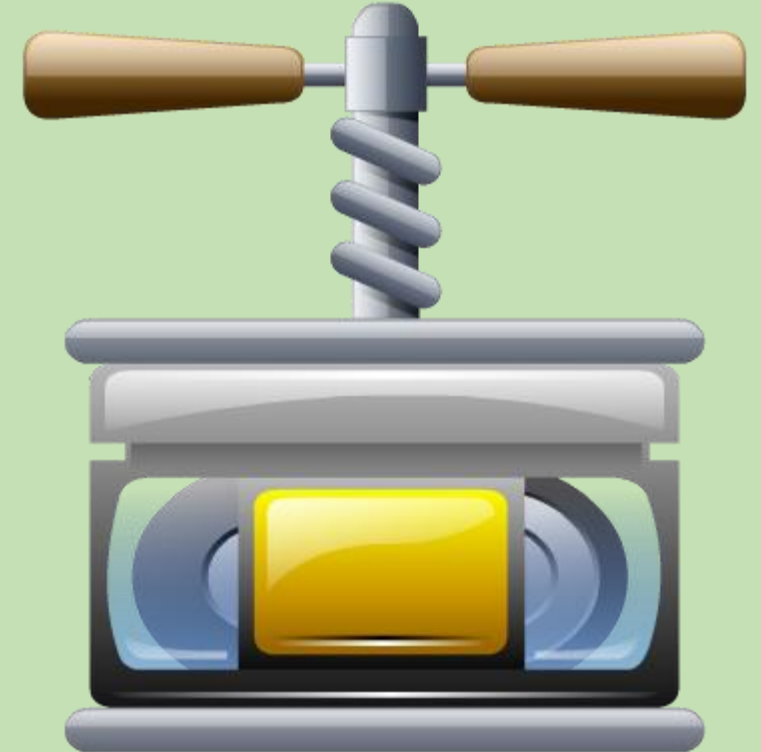


# Lesson 6

## File Compression

### Unit 1

## Data Representation



Mr. Teasdale

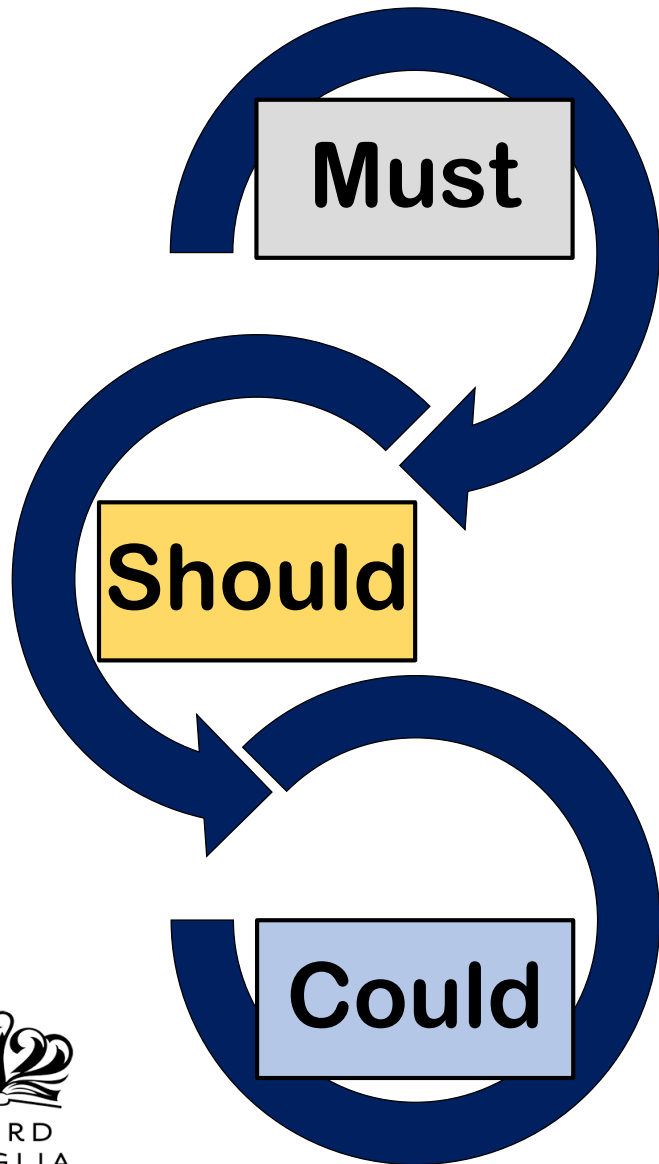
Today we are  
going to...



**Understand the principles of lossy  
and lossless file compression**



# Success Criteria



**Explain the importance of compressing files that are transmitted on the Internet**

SILVER

**Describe the difference between lossy and lossless compression**

GOLD

**Describe and apply the Run length encoding and Huffman coding algorithms to a problem**

PLATINUM



# Compression techniques

- Reduce file size
- Mostly used with sound, image and video file types
- Two types:
  - Lossy compression (JPG, GIF, MP3, MP4)
  - Lossless compression (PNG, TIF, SVG)
- Algorithms: **Run Length Encoding**, **Huffman coding**



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS



# Lossy compression

- Permanently removes some data
- Recreates the file using the remaining data and uses algorithms to guess the removed content
- Uncompressed data is not the same as the original
- Would this technique work for compressing a computer program?

```
# S CTIO 2
w ile gues = rr ctPas or :
    guess = i put "Try to guess t sswor ")
    g ses gue es + 1

rint("Pa swor uessed c r ctly
```



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

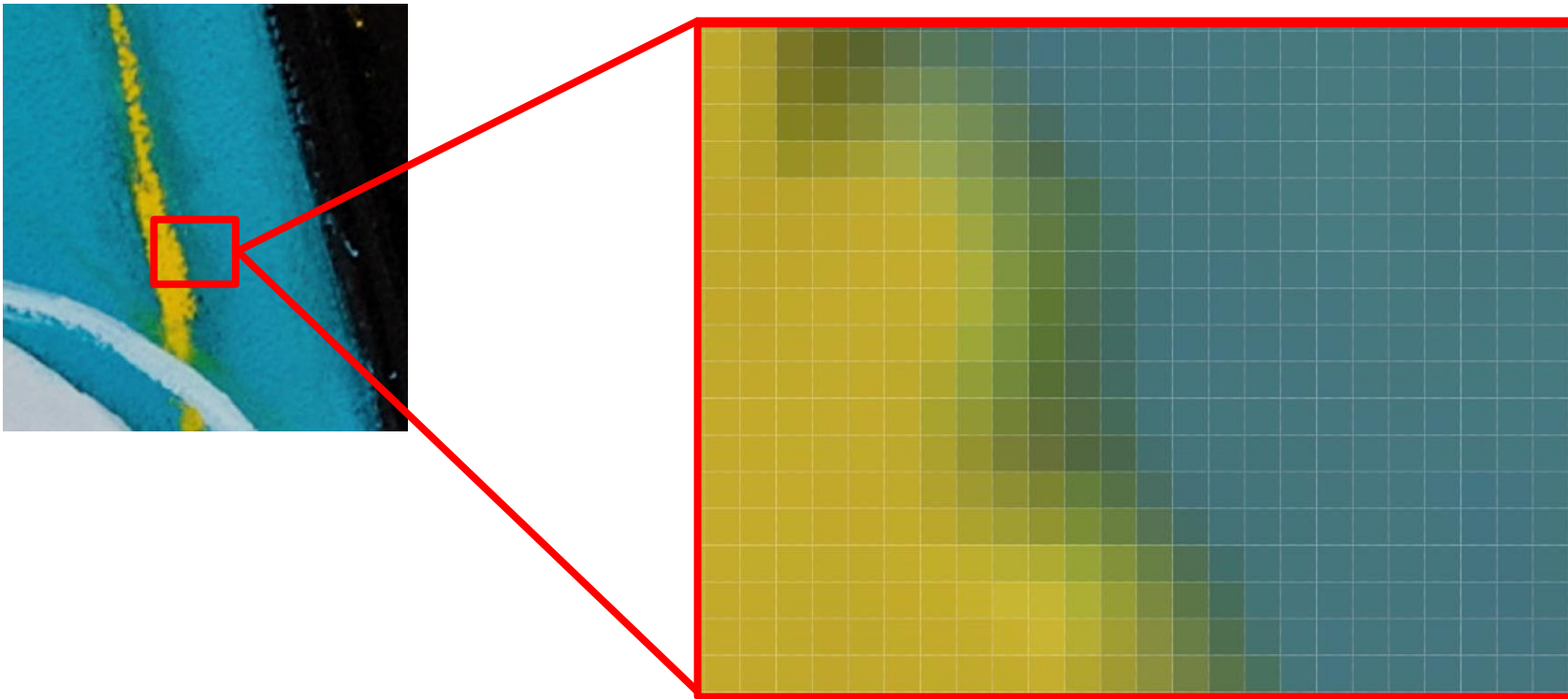
Y10/11  
IGCSE CS





# How lossy compression works

- Similarly coloured pixels are all made the same



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

### My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS







# Lossless image compression

- Finds groups of repeating data and records the data only once along with the number of times it was repeated



$$= 12 \times \text{[Yellow Square]} + 6 \times \text{[Teal Square]}$$

- When data is uncompressed it is restored exactly as it was in the original



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS







# Lossless text compression

- Finds patterns in the original text
- Encodes each pattern in a dictionary

*An eye for an eye,  
a tooth for a tooth*

.	0	0000
An_	1	0001
eye	2	0010
_for_	3	0011
an_	4	0100
,_	5	0101
a_	6	0110
tooth	7	0111



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS





# Lossless text compression

- 38 Characters including spaces = 38 bytes (assuming an 8-bit ASCII table is used)
- 48 bits = 12 bytes = 32% of original size (plus codes)

.	0	0000
An_	1	0001
eye	2	0010
_for_	3	0011
an_	4	0100
,_	5	0101
a_	6	0110
tooth	7	0111

1	2	3	4	2	5	6	7	3	6	7	0
0001	0010	0011	0100	0010	0101	0110	0111	0011	0110	0111	0000



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem





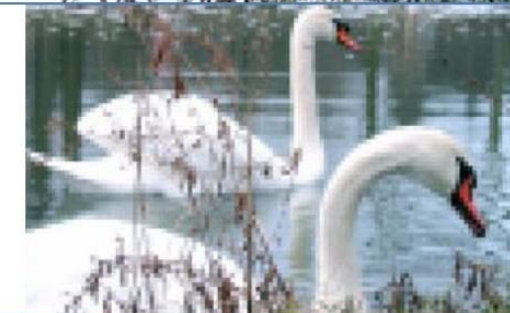
# Video files

- Compressing a video file reduces the resolution, dimensions and bit rate.
- Compressing a video file may also lead to poor quality and random coloured blocks on the screen. These blocks are called artefacts.
- MP4 and MOV are few examples of lossy video file formats.

Uncompressed video  
Duration 5 mins  
Size 50 MB



Compressed video  
Duration 5 mins  
Size 10 MB



11



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS





# Types of compressed files

12

Format	Type of compression	Application
PNG	lossless	Used for transferring images over the internet
JPEG	lossy	Higher compression rate than a PNG. Used in digital cameras and web pages.
GIF	lossless	Compresses images to a maximum of 8-bit depth. Not used for high quality images. A sequence of gif images is used to store animated graphics. Used for small images such as logos, icons, etc.
PDF	lossless	Encodes text and graphics



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS



# Task! Worksheet 6

- Complete **Task 1 – Lossless compression**

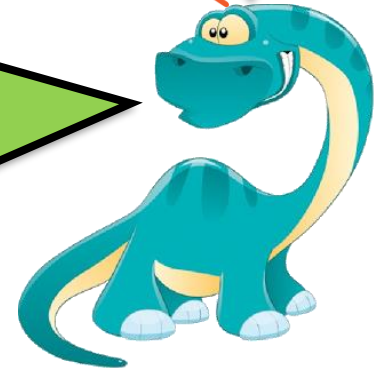


**Need help?**  
Checkout the  
revision notes in  
your notebooks



**Challenge?**

Checkout  
the images  
extension  
task!



## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem





# Transmission of data over IP

- Why use compression?
  - Download speeds are increased
  - Data allowances are reduced
  - Voice can be transmitted fast enough to keep up with speech



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem





# Downloading a music track

- Dancing Queen by ABBA = 3m 51sec = 231 seconds
  - MP3 quality = 128kbps
  - CD quality = 1411kbps
    - $231 \times 128\text{kbps} = 29,568\text{kbs} / 1024 / 8 = 3.6\text{MB}$
- OR:
- $231 \times 1411\text{kbps} = 325,941\text{kbs} / 1024 / 8 = 39.79\text{MB}$
  - 11.5 times faster with a compressed file
  - 36MB less download data used



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS





# Task! Worksheet 6

- Complete **Task 2**

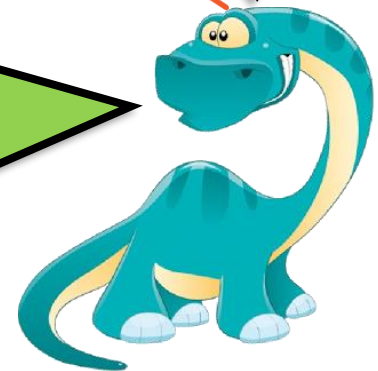


**Need help?**  
Checkout the  
revision notes in  
your notebooks



**Challenge?**

Checkout  
the images  
extension  
task!



## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem



# Run length encoding (RLE)

- is an example of a compression algorithm that converts the consecutive similar values into a code.
- This code consists of the identical value and the number of times this value is repeated.
- This is a lossless type of compression.

17



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

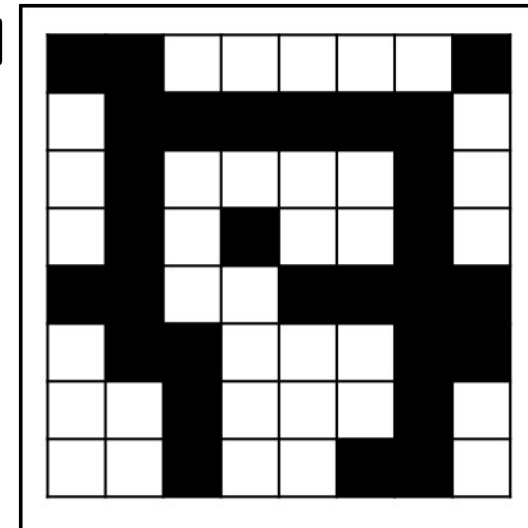
- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem





# Run length encoding (RLE)

- The computer stores binary value 1 for white and binary value 0 for black for each row of the image.
- The first row in the image can be represented as 2 0 5 1 1 0. This code represents 2 black pixels, 5 white pixels and 1 black pixels.
- Similarly, the second row in the image is represented as 1 1 6 0 1 1.



18



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

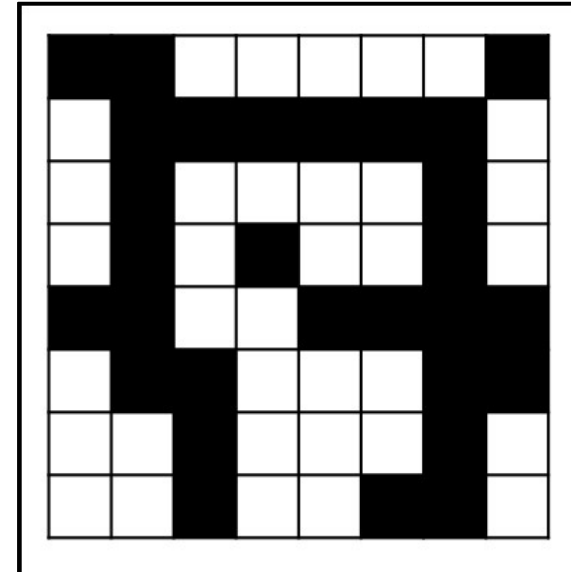
Y10/11  
IGCSE CS





# Run length encoding (RLE)

- This type of coding is not efficient if the file does not have many **runs**.
- In such cases, the file size may **increase** instead of getting compressed.
- RLE is thus used only in simple images with a **large area of same** colour.



19



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

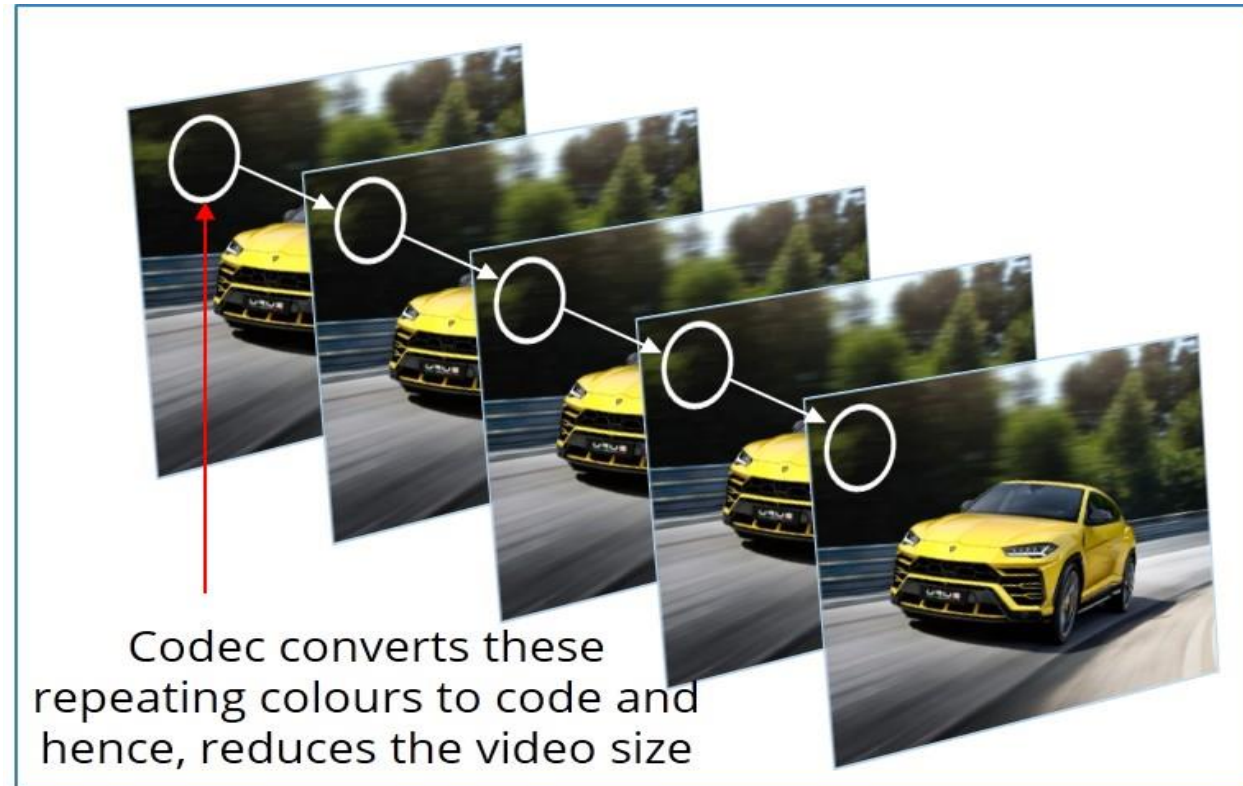
Y10/11  
IGCSE CS





# Run length encoding (RLE)

- RLE are also used to compress video files.



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem







# Huffman coding algorithm

- A compression technique used to reduce the number of bits that represent each letter.
- A binary tree is used to encode letters.
- A binary tree is a data structure made of nodes and is constructed based on hierarchy. A parent node in a binary tree has up to two child nodes.

21



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS





# Huffman coding algorithm

- In ASCII coding, each letter is represented using 7 bits.
- In Huffman coding, each letter is represented with a different number of bits.
- The most frequently appearing letters are represented with a lesser number of bits.
- The number of bits required to store information is reduced.



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS







# Huffman coding algorithm

- Consider the sentence: Betty ate butter.
- The frequency of characters in this sentence are shown in the table.
- There are 17 characters in total (including spaces).
- Therefore, the total number of bits used to represent their ASCII codes are:  $17 \times 7 = 119$  bits.

Letter	A	B	E	R	T	U	Y	Space
Frequency	1	2	3	1	5	1	1	3



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

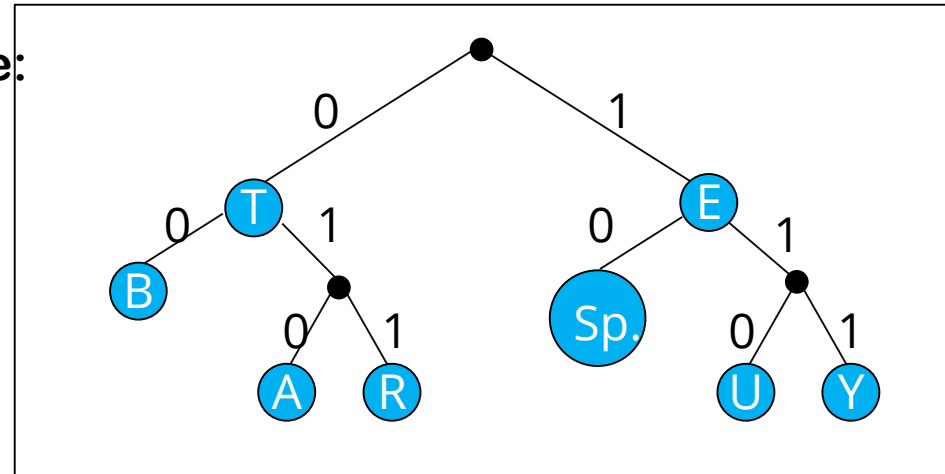
- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem





# How to do apply it?

- Consider the sentence: Betty ate butter.
- Each letter is now assigned a binary value:



Letter	A	B	E	R	T	U	Y	Space
Frequency	1	2	3	1	5	1	1	2
Binary value	010	00	1	011	0	110	111	10



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS





# How to apply it?

- Substituting these values in the sentence and calculating the total number of bits:  $3 + 4 + 3 + 3 + 5 + 3 + 3 + 4 = 28$  bits.
- Using Huffman coding, we have saved  $119 - 28 = 91$  bits.

Letter	A	B	E	R	T	U	Y	Space
Frequency	1	2	3	1	5	1	1	2
Binary value	010	00	1	011	0	110	111	10



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS

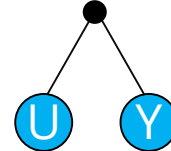
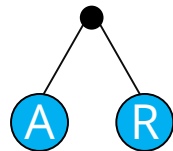




# How to apply it?

- We need the lowest number of bits for letters with higher frequency.
- Hence, we must award a slightly greater number of bits for letters with lesser frequency.
- We shall start with the letters with the lower frequency. Letters A, R, U and Y have the lowest frequency of 1.

Step  
1:



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

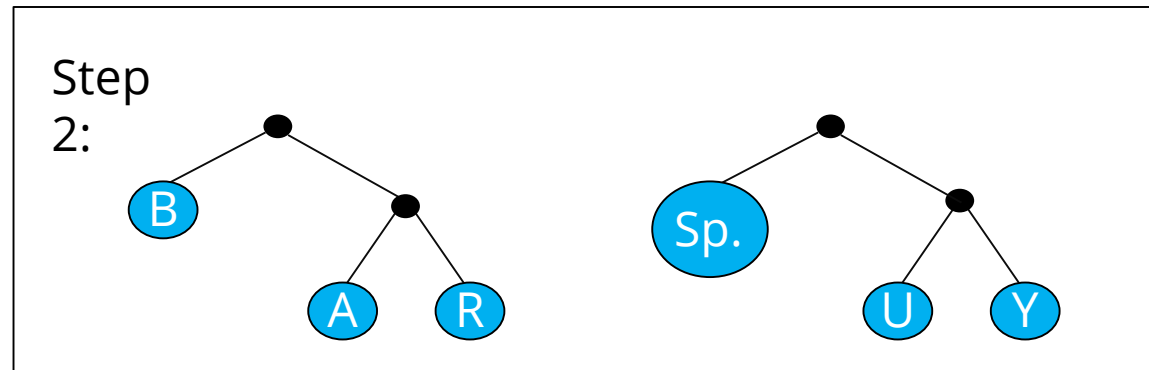
Y10/11  
IGCSE CS





## How to apply it?

- Next, let us consider characters B and space with a frequency of 2 each.



### Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

### My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

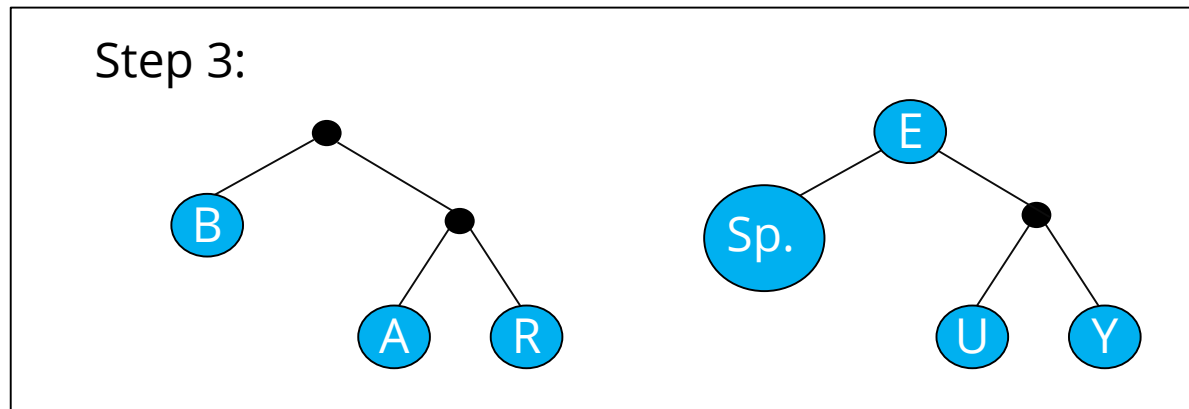
Y10/11  
IGCSE CS





## How to apply it?

- Next, let us consider letter E with a frequency of 3.



### Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

### My Success Criteria

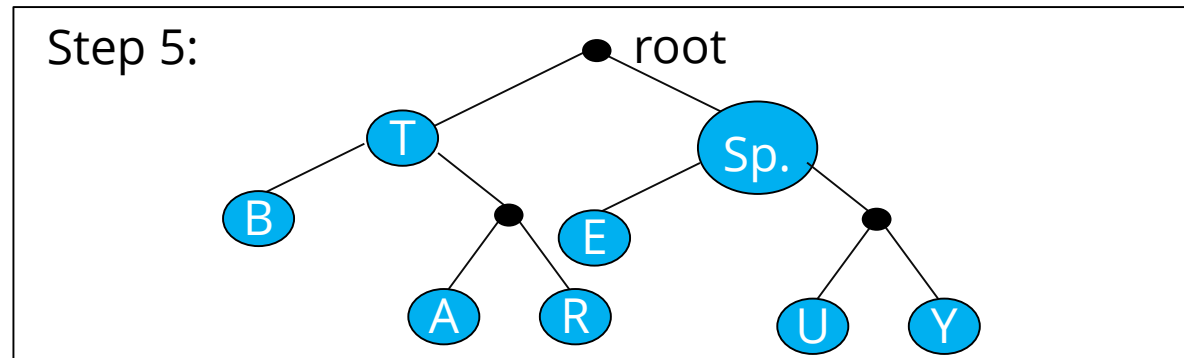
- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem





## How to apply it?

- The letter T has the highest frequency of 5. Let us include T and connect the nodes to the root.



### Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

### My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS

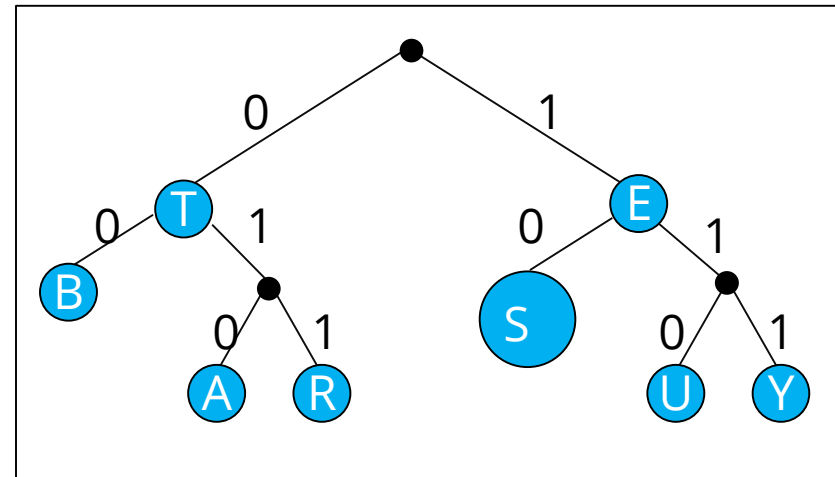






# How to apply it?

- The final step of forming a Huffman tree is giving binary values for each connection.
- Left branches are assigned a value of 0 and right branches are assigned a value of 1.
- Each path terminates to a leaf.



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS

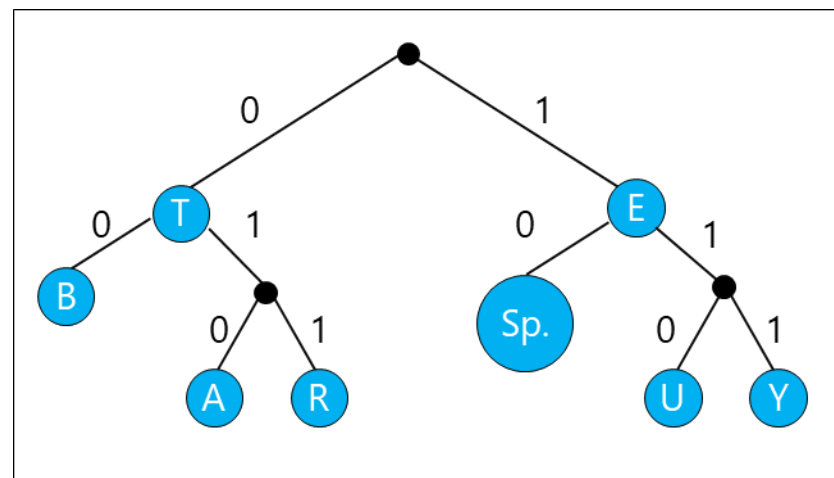




# How to apply it?

- Using the tree, each letter is assigned a binary value, starting from the root to the leaf.
- This is an example for using Huffman coding. There shall be different character coding for a letter.

Letter	A	B	E	R	T	U	Y	Space
Binary value	010	00	1	011	0	110	111	10



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS





# Benefits of compression

- **Smaller files = fewer packets = faster transmission**
  - Quicker to complete
  - Reduces traffic over the Internet
  - Less chance of collisions or transmission errors
- Improves download speed of video, sound (including speech used for VOIP systems) and image files
- Speeds up download of webpages that use images
- Reduces space on disk / servers
- Enables better streaming of music and video



## Literacy Focus

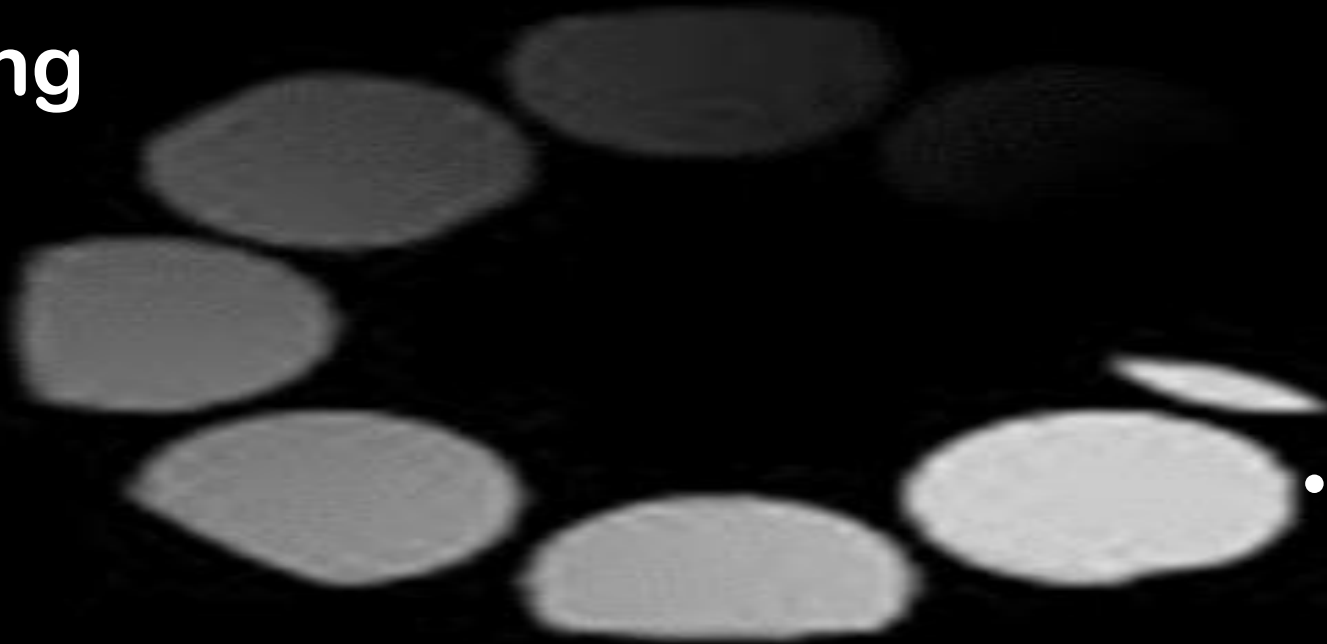
- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem



## Buffering



- Video or music streaming causes buffering if the download speed is slower than the playback speed
- How could you reduce the chances of people experiencing buffering if you were the website owner?

# loading



# Common file standards

- PDF – Fixed layout document that maintains its original appearance regardless of the software used to view it
- JPG – Lossy compressed files commonly used for website images
- GIF – Lossy compressed images using only 256 colours, used for simple web graphics
- PNG – Lossless compressed format using an alpha channel to preserve transparent backgrounds
- MP3 / MP4 – Lossy music / video formats



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS



# Task! Worksheet 6

- Complete **Task 3 – Compression uses**

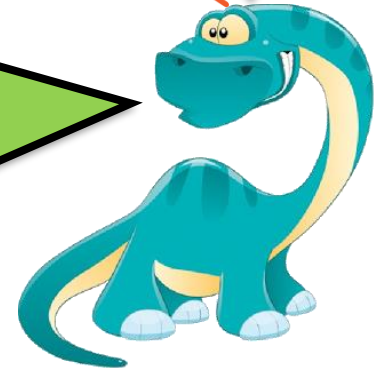


**Need help?**  
Checkout the  
revision notes in  
your notebooks



**Challenge?**

Checkout  
the images  
extension  
task!



## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem



# Checkpoint



✓ How confident are you?



**Must**

Explain the importance of compressing files that are transmitted on the Internet



**Should**

Describe the difference between lossy and lossless compression



**Could**

Describe and apply the Run length encoding and Huffman coding algorithms to a problem







# Cool down

1. Complete the skills 'checklist'
2. Answer the confidence question



3. Reflect on your learning

4. Review..... Quizlet Live

<https://quizlet.com/520861395/live>



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

### My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS



# Let's Review



**Must**

Explain the importance of compressing files that are transmitted on the Internet

SILVER

**Should**

Describe the difference between lossy and lossless compression

GOLD

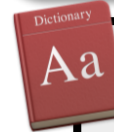
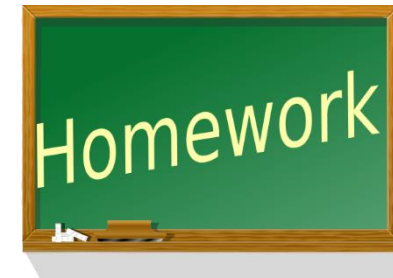
**Could**

Describe and apply the Run length encoding and Huffman coding algorithms to a problem

PLATINUM

# Homework

Homework is in  
your notebooks,  
complete for next  
lesson!



## Literacy Focus

- ☐ Bitmap
- ☐ Colour depth
- ☐ Metadata
- ☐ Pixel
- ☐ Resolution
- ☐ Vector

## My Success Criteria

- ☐ Explain the importance of compressing files that are transmitted on the Internet
- ☐ Describe the difference between lossy and lossless compression
- ☐ Describe and apply the Run length encoding and Huffman coding algorithms to a problem

Y10/11  
IGCSE CS

