Your notes

GCSE Computer Science: revision notes

Data compression

Contents

1. Why compress files?
2. Lossy compression
3. Lossless compression
4. Run-length encoding (RLE)
5. Huffman coding (AQA)
6. Key terms
7. Quick questions
8. What to do next

Why compress files?

- **Compression** makes files smaller, so they take up less storage space and are quicker to send or download, using less **bandwidth**.

Lossy compression

- **Lossy compression** permanently removes some data, usually detail that people are unlikely to notice. It greatly reduces the file size, but the original can't be recovered exactly.
- It is used for photos (JPEG), music (MP3) and video. It is not suitable for text or program files.

Lossless compression

- **Lossless compression** makes a file smaller without losing any data, so the original can be recovered exactly.
- It is used for text, program files and some images (PNG). It usually reduces the file size less than **lossy compression**.

Run-length encoding (RLE)

- RLE stores each run of repeated data as a count and a value: AAAAABBB becomes 5A 3B. In binary, `0000011100` becomes 5 `0`s, 3 1s and 2 `0`s.
- It works well when there are long runs of repeated values, but can make data bigger when there are few repeats.

Huffman coding (AQA)

- **Huffman coding** gives shorter binary codes to the characters that appear most often, using a Huffman tree. It is lossless.

- Bits needed = the total of (frequency × code length) for each character. A text with 4 As, 2 Bs and 1 C, coded A = 0, B = 10 and C = 11, needs 4 × 1 + 2 × 2 + 1 × 2 = 10 bits, compared with 7 × 7 = 49 bits in 7-bit ASCII.

Your notes

Key terms

Compression	Making a file smaller.
Lossy compression	Compression that permanently removes some data.
Lossless compression	Compression that keeps all the data, so the original can be recovered exactly.
Run-length encoding	Storing runs of repeated data as a count and a value.
Huffman coding	Lossless compression that gives shorter codes to more frequent characters. (AQA)
Bandwidth	The amount of data that can be sent over a network each second.

Quick questions

Cover the answers and test yourself: remembering is what makes it stick.

1. Why are files compressed?
2. Which type of compression permanently removes data?
3. Which type of compression should be used for a program file?
4. Why is lossy compression not suitable for text?
5. Which file type usually uses lossy compression?
6. What is AAAAABBB in run-length encoding?
7. When does run-length encoding work well?
8. What does lossless compression allow?

Answers

1. To make them smaller, so they use less storage and transfer faster
2. Lossy
3. Lossless
4. Removing data would change the text
5. MP3
6. 5A 3B
7. When there are long runs of repeated values
8. The original file to be recovered exactly

What to do next

Practise 12 questions on this topic

Try 5 exam questions, with mark schemes

Read the lesson, one step at a time

Your notes