Definitions and Concepts for AQA Computer Science AS-level

Topic 5: Fundamentals of Data Representation

5.1 Number Systems

5.1.1 Natural Numbers

Natural Numbers (ℕ): The set of positive integers and 0. They can be used as cardinal(counting) or ordinal(ordering) numbers. ℕ = {0, 1, 2, 3, ...}

5.1.2 Integer Numbers

Integers (ℤ): The set of numbers with no fractional part. The natural numbers are a subset of the integers. ℤ = {..., -3, -2, -1, 0, 1, 2, 3, ...}

5.1.3 Rational Numbers

Rational Numbers (ℚ): The set of numbers that can be expressed as the ratio of two integers. The integers are a subset of the rational numbers since all integers can be expressed as a ratio with 1. ℚ = {0, ½, 0.75, 0.111111…, 300.5, -42, ...}

5.1.4 Irrational Numbers

Irrational Numbers: Number which cannot be expressed as a ratio of two integers, and hence do not lie within the set of rational numbers. {π, $\sqrt{2}$, e, ...}

5.1.5 Real Numbers

Real Numbers (ℝ): The set of numbers that can represent real world quantities and have an imaginary part of 0. Rational and irrational numbers are all members of the real numbers. ℝ ={π, 1.5, -7, ¾, 2, 100000000, -11.3432, ...}

5.1.6 Ordinal Numbers

Ordinal Numbers: Natural numbers used to describe numerical position or order of objects.

5.2 Number Bases

5.2.1 Number Base

Binary: A number system that only uses ones and zeros to represent numbers (a base 2 system).

Decimal: A number system that only uses 10 characters (0 to 9) to represent numbers (a base 10 system).

Hexadecimal: A number system that only uses 16 characters (0 to 9 and A to F) to represent numbers (a base 16 system).

Number Base: The number of unique digits used by a particular number system to represent numbers.

5.3 Units of Information

5.3.1 Bits and Bytes

Bit: A binary digit used by computers as the fundamental unit of information.

Byte: A group of 8 bits.

5.3.2 Units

Binary Prefix: A prefix to a unit representing a power of 2. (kibi=2^{10}, mebi=2^{20}, gibi=2^{30}, tebi=2^{40})

Decimal Prefix: A prefix to a unit representing a power of 10. (kilo=10^{3}, mega=10^{6}, giga=10^{9}, tera=10^{12})

5.4 Binary Number System

5.4.1 Unsigned Binary

Signed Binary: A binary number system that can represent both positive and negative numbers.

Unsigned Binary: A binary number system that can only represent positive numbers.

5.4.3 Signed Binary Using Two's Complement

Two's Complement: A coding scheme used in signed binary to represent negative as well as positive numbers. A negative number is represented by flipping all its digits and adding 1 to the most significant bit.

5.4.4 Numbers with a Fractional Part

Fixed Point Form: A form used to represent numbers with a fractional part in any number system. Digits after the fixed point are multiplied by the base raised to a negative power.

5.5 Information Coding Systems

5.5.1 Character Form of a Decimal Digit

Character Code: A unique binary representation of a character. Not to be confused with the binary representation of a decimal digit, which is it's numerical value in the binary system.

5.5.2 ASCII and Unicode

ASCII: A character set used to represent alphanumeric characters or symbols as a set of 8 bits.

Unicode: A character set that is a superset of ASCII. It is used to represent alphanumeric characters or symbols as an integer code point which is equal to the character's ASCII code.

5.5.3 Error Checking and Correction

Check Digits: A method of checking codes for errors during data transmission by adding an extra digit to the end (usually calculated/processed from digits in the code itself) that checks whether the data is accurate.

Majority Voting: A method of checking binary codes for errors during data transmission by sending each bit multiple times, in a set. The receiver takes the value with most occurrences in a set as the value for that bit.

Parity Bits: A method of checking binary codes for errors during data transmission by counting the number of ones and zeroes present.

5.6 Representing Images, Sound and Other Data

5.6.2 Analogue and Digital

Analogue Data: Data whose values can vary continuously and take on any value between two extremes.

Analogue Signals: A transmission of a set of analogue data structures, that varies with time, between computational processes.

Digital Data: Data whose values can vary discretely and can only take on one of a finite number of values between two extremes.

Digital Signals: A transmission of a set of digital data structures, that varies with time, between computational processes.

5.6.3 Analogue/Digital Conversion

Analogue to Digital Converter (ADC): An integrated circuit capable of converting continuous analogue data to discrete digital data for a computer.

Digital to Analogue Converter (DAC): An integrated circuit capable of converting discrete digital data from a computer to continuous analogue data.

5.6.4 Bitmapped Graphics

Bitmapped Graphics: An image composed of an array of pixels each with an allocated number of bits, arranged to form an image. Also known as raster graphics.

Bitmap Storage Requirements: The amount of storage required for a bitmapped image is at least its (image size) × (colour depth).

Colour Depth: A measure of the amount of colour used in an image, expressed in terms of the number of bits per pixel.

Image Size: The total number of pixels in an image expressed in terms of its dimensions: (width in pixels) × (height in pixels).

Metadata: Data related to the image file data itself. This includes image properties such as width, height and colour depth.

Resolution: A measure of the total number of pixels in an image, typically expressed in terms of the number of dots/pixels per inch.

5.6.5 Digital Representation of Sound

Nyquist theorem: A sufficiently accurate digital waveform of an analogue signal would require a sampling rate of at least twice the highest frequency that appears in the original analogue signal.

Sample Resolution: The number of bits used to represent a single sample.

Sampling Rate: The number of samples taken per second.

Sound Sampling: The process of converting analogue sound waves to a digital waveform, by storing a finite number of readings in binary.

5.6.6 Musical Instrument Digital Interface (MIDI)

Event Messages: Binary data transmitted between the MIDI device and computer processor that carries properties controlling when and how sounds are produced.

MIDI: Musical Instrument Digital Interface is a protocol for ADC audio transmission to a digital

interface used for the majority of electronic musical instruments and computers.

5.6.7 Data Compression

Dictionary-based Coding: A type of lossless compression where text is searched for entries that match the entries in a dictionary. Entries are substituted by a unique code which can then be translated.

Lossless Compression: A compression algorithm that retains all the data in the file by only storing the instructions needed to reconstruct the original file. No data is lost.

Lossy Compression: A compression algorithm that removes non-essential data from a file leading to a noticeable decrease in accuracy of the data. Data lost is non-recoverable.

Run-Length Encoding: A type of lossless compression where repeated occurrences of the same data (like several pixels of the same colour in an image) are stored as single data values with their counts.

5.6.8 Encryption

Encryption: The process of converting the original data (plaintext) into a form which cannot be understood by unauthorised users (ciphertext) using an encryption algorithm (cipher).

Caesar Cipher: A substitution cipher where each letter of plaintext is substituted for another that is a fixed number of letters ahead in the alphabet, which becomes the ciphertext.

Vernam Cipher: A cipher that uses a one-time pad (a secret random key) to convert each character to cipher text by modularly adding it with the corresponding character of the key. This is impossible to decrypt without a key.