Questions matter

Subject specific vocabulary

The following subject specific vocabulary provides definitions of key terms used in our GCSE Computer Science (8525) specification.

Your students should be familiar with, and gain understanding from all these terms.

3.1 Fundamentals of algorithms

Algorithm
A sequence of steps that can be followed to complete a task.

Abstraction
The process of removing unnecessary details from a problem.

Decomposition
Breaking a problem into a number of subproblems, so that each subproblem accomplishes an identifiable task, which might itself be further subdivided.

Linear search
An algorithm that sequentially checks each element in an array or list until a match is found or the end is reached.

Binary search
An algorithm that compares a search item to the middle item of an array or list. If they do not match, half of the array or list can be eliminated and the search continues in the other half. The array or list needs to be sorted so that a binary search can be performed.

Merge sort
The merge sort is a 'divide-and-conquer' algorithm. The original array or list is broken down into smaller and smaller parts until we have individual elements.
The merge sort continues by building back sorted groups until we end with the full sorted result.

Bubble sort
A bubble sort works by making passes through the array or list.
Whilst making a pass, we compare neighbours and swap them if they are out of order.
After a series of passes, the array or list will be sorted.

3.2 Programming

Data type
Each variable has a data type. The data type determines what kind of value the variable will hold and the set of operations that can be performed on it.

Integer
A whole number.

Real/Float
Numbers that can include fractions/values after the decimal point.

Boolean
A data type that can represent the two values True and False.

Character
Stores a single character in a fixed-length format related to a specific character set (for example ASCII).

String
A string is a sequence of characters.

Variable
An identifier that represents a memory location that can store a value. This value can change during execution of a program.

Variable declaration (examples)

C#
```
int Age = 17;
```
VB
```
Dim Age As Integer = 12
```
Python
```
Age = 12
```

Constant
An identifier that represents a memory location that can store a value. This value cannot change during execution of a program.
Note: Python does not have constants but convention is that an identifier in capitals should indicate the concept of a constant.

Constant declaration (examples)

C#
```
const int Months = 12;
```
VB
```
Const Months As Integer = 12
```
Python
```
MONTHS = 12
```

Array
A data structure that stores a collection of elements of the same data type under a single variable name and is accessed through an index.

Assignment
An assignment statement sets the value of a variable/constant.

Iteration
Iteration is the repetition of a block of code.

Definite iteration
Definite iteration is the repetition of a block of code for a set number of times.
An example structure is a `FOR` loop.

Indefinite iteration
Indefinite iteration is the repetition of a block of code being controlled by a condition.
The number of iterations to be performed is not known when entering the block of code.
Example structures are `WHILE` and `DO LOOP UNTIL`.

Selection
A selection statement allows a block of code to be executed only when a certain condition is satisfied.
Examples of selection statements include:
`IF THEN ELSE`
`SWITCH/CASE/MATCH`

Nested (selection/iteration)
When an iteration statement appears inside another iteration statement it is considered nested.
When a selection statement appears inside another selection statement it is considered nested.

Random number generation
The use of algorithms to produce sequences of numbers that appear unpredictable.

Subroutine
A named 'out of line' block of code that may be executed (called) by simply writing its name in a program statement. It can be reused multiple times within a program.

Parameters
Parameters are variables which define the data to be passed into a subroutine when the subroutine is called.

Local variables
Local variables:
- are declared within a subroutine
- exist only while a subroutine is executing
- are accessible only from within the subroutine.

Structured approach to programming
The structured approach to programming includes modularised programming, clear well-documented interfaces (local variables, parameters) and return values.

Validation routine
A routine that contains code to check the validity of data.
Validity of data refers to data being reasonable/plausible/acceptable. This might involve range checks, length checks and presence checks.

Authentication routine
A routine that contains code that verifies an identity. This might involve the entry of a username and password.

Testing
Testing is a process to help the programmer to remove bugs from software and to ensure that it works correctly.

Types of testing data:
- Normal (typical)
- Boundary (extreme)
- Erroneous

Syntax error
A syntax error is when code does not follow the rules of the language grammar.
A syntax error will prevent code from being executed.
These errors are caught by compilers and interpreters and can include misspelt keywords and missing symbols such as brackets and punctuation symbols.

Logical error
A logical error is a mistake in code that causes it to produce unexpected or incorrect responses. A logical error does not prevent a program from running.

3.3 Fundamentals of data representation

Decimal – base 10
The decimal number system (base 10) uses 10 different numerals 0,1,2,3,4,5,6,7,8,9

Binary – base 2
The binary number system (base 2) uses 2 different numerals 0,1

Hexadecimal – base 16
Hexadecimal is a number base that uses the numerals 0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F
It has a close link to binary as 4 bits can be represented by 1 hexadecimal digit.
Hexadecimal is a more compact form than binary and therefore can take up less space on a display and take less time to enter.

It Is also easier for people to understand/remember.
Note: hexadecimal does not take up less memory space (as a computer only stores binary)

Bit
The fundamental unit of information (a 1 or a 0).

Byte
A group of 8 bits.

Kilo	1000 bytes	(10^3 bytes)
Mega	1000 Kilo	(10^6 bytes)
Giga	1000 Mega	(10^9 bytes)
Tera	1000 Giga	(10^{12} bytes)

Logical binary shift
A logical binary shift involves moving the bits of a binary number to the left or right.
Vacated positions are filled by zeros.
A shift to the left corresponds to multiplying by 2.
A shift to the right corresponds to dividing by 2.

ASCII
ASCII is a character encoding system that uses a 7-bit notation that allows for the representation of 128 unique characters.

Unicode
Introduced to support a wider range of characters due to increased communication across countries.
One of the latest standards defines around 155000 characters.
It can be encoded in a variety of ways with up to 32 bits per character.

Character set
A collection of characters (letters, numbers, symbols, etc.) that a computer system can process and use to represent text, with each character assigned a unique numerical code.

Pixel
The term pixel stands for picture element. A pixel is a single point of colour in an image.

Bitmap
A bitmap is where images are broken down into a rectangular grid of pixels.
Each pixel will have its own colour value.

Analogue
Continuous data – can take any value in a given range.
Between any two analogue values there is another value.

Digital
Digital data has discrete values.
There are gaps (jumps) between digital data values.

ADC – analogue to digital convertor
An ADC takes samples of the analogue signal at regular intervals.
The amplitude of each sample is approximated to an integer value.
Each sample is encoded as a binary value.

Sampling rate
A measure of the number of samples taken per second from an analogue signal to create a digital signal.

Sample resolution
A measure of how many bits are used to represent each sample.

Data compression
The process of reducing the storage space required for data by encoding the data using fewer bits than the original representation.

Huffman coding
Huffman coding is a lossless method of data compression.
Huffman coding produces variable-length codes where the length of the code depends on the frequency with which a symbol appears in the original data.

Run-length encoding (RLE)
Run-length encoding is a lossless method of data compression.
Sequences of identical data values are replaced with a single data value and the count of its repetitions.

3.4 Computer systems

Hardware
The physical/electronic components of the computer system.

Software
Program/code/instructions executed by the hardware.

Truth table
A representation of the output of a logical expression or circuit for all possible combinations of input variables.

Logic gate
An electronic component that takes one or more binary inputs and performs a basic logical operation (such as AND, OR and NOT) to produce a single binary output.

System software

System software is software that manages the computer system resources and acts as a platform to run application software.

Application software

Application software is software that performs end-user tasks. Tasks that the user may wish to complete even without a computer (e.g. write a letter).

Utility software

Utility software is software that can help maintain a system.
Examples include:

- backup software
- encryption software
- compression software.

Operating system (OS)

The role of an operating system is to hide the complexities of the hardware from the user.
The operating system handles:

- resource management (processor(s), memory, input/output (I/O) devices)
- applications
- security.

Low-level language

A low level language is one that is closer to the hardware and the computer's instruction set.
Machine code and assembly language are examples of low-level languages.

Machine code

Machine code is another name for the actual binary instructions that can be executed by a processor and is specific to a processor or family of processors.

Assembly language

Assembly language uses mnemonics that represent machine code instructions.
Assembly language needs translating (assembling) before it can be run.

High-level language

A type of language that is more abstracted from the hardware of the computer and in which instructions are more powerful than in a low-level language.
HLLs use English-like keywords.
HLLs support structured statements – such as iteration/selection.
HLLs support use of local variables and parameters.

Program translator

A processor can only execute machine code instructions. So anything else (assembly language, HLL code) needs to be translated first into machine code before it can be run.
Assemblers, interpreters and compilers are examples of translators.

Interpreter

An interpreter analyses the source code line by line.
An interpreter will run the source code up to the first error.
Interpreted code will execute more slowly than executing the code produced by a compiler.
An interpreter always needs a copy of the source code at runtime.
Interpreted code is more portable.

Compiler

A compiler produces object code from the source code.
A compiler translates the whole source code at once.
A compiler will not produce an executable file (will not compile) if an error is encountered.
Compiled code can only be executed on a machine with the same instruction set.
Once compiled, the object code (executable) can be executed without the need to have the compiler.

Assembler

An assembler translates assembly language into machine code.

Central Processing Unit (CPU)

The primary component responsible for processing data and executing instructions.

Arithmetic logic unit

A part of the processor where arithmetic and logical operations are performed.

Control unit

A part of the processor which

- decodes instructions
- coordinates/directs the execution of instructions
- sends control signals to other components.

Clock

Synchronises operations of the computer by sending regular timing signals to the processor.

Register

A memory location on a processor that can be accessed quickly.
A register can hold data used when executing an instruction and the result of executing an instruction.

Bus

A bus is a series of lines (wires) that transfers data, instructions or memory addresses between components inside a computer.

Clock speed

The clock sends out regular pulses to coordinate the components of the computer.
The frequency of these pulses Is known as the clock speed and measured in hertz, where 1 hertz = 1 pulse per second.

Core
A processing unit within a processor is known as a core.
Each core can fetch, decode and execute instructions independently of the other cores.

Cache
Cache is a small area of fast memory built into a modern processor.
Cache holds frequently used instructions and data that the processor is likely to reuse.

Volatile/non-volatile
In terms of storage devices, volatility refers to losing data contents when power is removed.
A volatile storage device does not retain data contents when power is removed.
A non-volatile storage device will retain data contents even when power is removed.

RAM – Random Access Memory
Main memory is considered to be any form of memory that is directly accessible by the CPU (except for cache and registers).
Random Access Memory is volatile main memory.
It is used to hold currently running programs and associated data.
It can be read from and written to.
Its contents are lost when power is removed.

ROM – Read Only Memory
Read Only Memory is non-volatile main memory.
It is used to hold instructions that are needed for the computer to run especially when the computer is turned on.
It can be read from but not written to.
Its contents are kept when power is removed.

Secondary storage
Secondary storage is considered to be any non-volatile storage mechanism not directly accessible by the CPU.

Solid state
Solid state devices use non-volatile memory to store data. They have fast data access and write speeds and are durable as they do not have any moving parts.

Magnetic
Magnetic devices such as hard-disk drives have a spinning disk and use magnetic fields to represent 0s and 1s.

Cloud storage
Storing data at a remote location, using magnetic and/or solid state storage, and allowing online access is known as cloud storage.
Access to cloud storage is made via the Internet.

Embedded system
An embedded system is a computer system built into a larger device.
An example of an embedded system is the control system for a microwave or a washing machine.

3.5 Fundamentals of computer networks

Computer network
A group of devices connected together so that they can communicate.

PAN – Personal Area Network
Devices connected and communicating over a small distance around one person.

LAN – Local Area Network
Devices connected and communicating over a small geographical area.
They are often owned and controlled/managed by a single person or organisation.

WAN – Wide Area Network
Devices connected and communicating over a large geographical area.
They are often under collective or distributed ownership.

Network protocol
A set of rules that allow devices/networks to communicate/transfer data.

TCP – Transmission Control Protocol
TCP is a transport layer protocol.
It is used to establish communication between two end points and ensure that this is error free.

IP – Internet Protocol
IP is an internet layer protocol.
IP is responsible for routing.
An IP address is given to a device when it connects to a network.
The IP address consists of four bytes of information and is typically represented as a dotted quad with each having a value in the range 0 to 255.
An example is:
`10.10.100.4`

HTTP/HTTPS
HyperText Transfer Protocol
HyperText Transfer Protocol Secure
An application layer protocol concerned with sending and receiving web pages between a webserver and a client (browser). HTTPS adds encryption to the protocol allowing for secure data movement.

SMTP
Simple Mail Transfer Protocol
An application layer protocol.
To send emails to another email server/client.

IMAP
Internet Message Access Protocol
An application layer protocol.
To receive emails from an email server.
To allow management of emails on the server.

Authentication
Authentication is the process of confirming someone is who they say they are.
This can be completed by a variety of methods including username and password.

Encryption
Using an algorithm to convert a message into a form that is not understandable without the key to decrypt it.

MAC address (Media Access Control address)
A unique identifier assigned to a network device.

Firewall
A firewall is a network security device that monitors incoming and outgoing network traffic and decides whether to allow or block specific traffic based on a defined set of security rules.
A firewall generally has a set of rules that will either allow or deny traffic and these can be set to look at:
- IP address (sender/destination)
- Ports
- MAC address (sender/destination)

For example, a rule could block all incoming FTP traffic on ports 20 and 21.

MAC address filtering
MAC address filtering allows devices to access, or be blocked from accessing, a network based on the physical address embedded within the device's network adapter.

TCP/IP model
The TCP/IP model is a protocol stack used when connecting a device to the Internet.
It is made up of four layers: Application, Transport, Internet and Link.

Application layer
The application layer controls how data moves in and out of applications from the transport layer. Applications include email clients, web browsers and protocols such as HTTP, HTTPS and IMAP.

Transport layer
The transport layer breaks down messages into smaller chunks (packets).
This layer sets up the communication between the two hosts and they agree settings such as the size of packets.
Each packet is given a sequence number so the receiver can put the message back together again. The transport layer uses protocols such as TCP.

Internet layer
The Internet layer adds the source and destination IP addresses to the packet and packages the data for transmission.
This layer is concerned with the routing of packets across the Internet.

Link layer
The link layer includes the network interface card and drivers necessary to send data out onto the network. The link layer is where the network hardware such as the NIC (network interface card) is located. OS device drivers also operate here.

3.6 Cyber security

Cyber security
Cyber security consists of the processes, practices and technologies designed to protect networks, computers, programs and data from attack, damage or unauthorised access.

Social engineering
Social engineering is the art of manipulating people so they give up confidential information.

Blagging
Blagging is the act of creating and using an invented scenario to engage a targeted victim in a manner that increases the chance the victim will divulge information or perform actions that would be unlikely in ordinary circumstances.

Phishing
Phishing is a technique of fraudulently obtaining private information, often using email or SMS.

Shouldering
Shouldering is observing a person's private information over their shoulder e.g. cashpoint machine PIN numbers.

Malware
Malware is an umbrella term used to refer to a variety of forms of hostile or intrusive software.

Virus
A small program, with a malicious purpose, attached to another program or data file.
It replicates itself by attaching itself to other programs.

Trojan
A program, with a malicious purpose, that hides in or masquerades as desirable software, such as a utility or a game, but attacks computers it infects.

Spyware
Malicious software that gathers data from a user and sends it back to a third-party without consent of the user.

Pharming
Pharming is a cyber attack intended to redirect a website's traffic to a fake website.

Weak password
A weak password is one that is easy for a human or computer to guess.

Default password
A default password is usually pre-configured by a manufacturer with the assumption that it will be changed by the user. Default passwords are often well known or can be easily discovered.

Misconfigured access rights
Occurs when users or groups are given the wrong permissions to access resources.

Unpatched software
Unpatched software is software with known security vulnerabilities that have not been fixed through applying a software update.

Removable media
Any removable secondary storage device onto which data can be copied and easily removed from the computer, e.g. USB memory stick.

Penetration testing
Penetration testing is the process of attempting to gain access to resources without knowledge of usernames, passwords and other normal means of access.

Biometric measures
Unique physical characteristics that can be used to identify individuals, for example, fingerprints, voice, facial scans.

CAPTCHA
Completely automated public Turing test to tell computers and humans apart.
A type of challenge request that is easy for humans to solve but hard for computers.

3.7 Relational databases and structured query language (SQL)

Database
An organised collection of persistent data.

Relational database
A database with multiple tables that are linked together/that reference each other.

Table
A structured collection of data organised into records (rows) and fields (columns).

Record
A single entry in a table. It represents the whole information about a particular entity (e.g. a student, a teacher).

Field
A single piece of data within a record (e.g. a first name, last name).

Primary key
A field that is unique for every record in a table.

Foreign key
A field in one table that is also a primary key in another table.
Foreign keys are used to link one table to another.

Data inconsistency
When the same piece of data is stored in various parts of the database, but instances of the data are not identical leading to unreliable information.

Data redundancy
The unnecessary duplication of data within a database.

SELECT
A select query is used to retrieve data from a relational database. It uses the following command structure:

```
SELECT field1, field2, …
FROM table1, table2, …
WHERE criteria
ORDER BY field1 ASC|DESC
```

INSERT

An insert query is used to insert data into a relational database. It uses the following command structure:

```
INSERT INTO table (field1, field2, …) VALUES (value1, value2, …)
```

UPDATE

An update query is used to edit or update the contents of a record in a relational database. It uses the following command structure:

```
UPDATE table SET field1=value1, field2=value2 …. WHERE criteria
```

DELETE

A delete query is used to delete data from a relational database. It uses the following command structure:

```
DELETE FROM table WHERE criteria
```

Version 1.1
June 2025