Information, Data, Data units and Units of Information

Computer Science

Computer science (or *computing science*) is the study of the information and computation and their implementation and application in computer systems.

Data

Data are representations of facts in a form suitable for processing by computers ("symbols").

Data on its own carries no meaning.

Analog Data

- continuous series of values

Digital Data

- represented by discrete values - usually zeros and ones

Main forms of data

- *graphical* – pictograms, pictures, photo, video – visual signals
- *audio* – sound signals
- *numerical* – digits – representations of quanta
- *textual* – letters, words – representations of words - purposefully written or transcribed from speech

Information

There is no universal definition of what the information is – it depends on the context. It provides answers to "who", "what", "where", and "when" questions.

Information is data that has been **processed** in such a way as to be **meaningful** to the person/system that receives it. **It increases understanding or decreases uncertainty.**

Information = data + context

Context = how to interpret the data (date, physical quantity, age, elevation …)

Information has two significant attributes:

- it is **measurable** – there are ways how to quantify the information.
- it has its **receiver that understands it.**

For example, a red traffic light is data. The meaning that we attach to this data is "Stop". When you drive up to a red light and stop, you do so because your brain sees the data and processes it.

The fact the data is not equal to information can be simply demonstrated by an ***encrypted*** *message – the bits can be captured, however they do not bring any information until it is decrypted.*

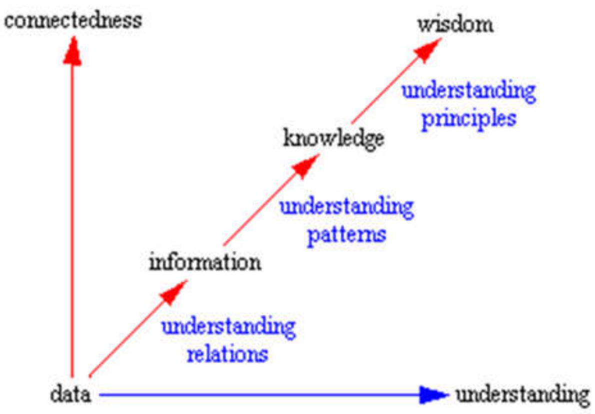
Knowledge

- application of data and information; answers "how" questions

Wisdom

- cannot be easily defined – might be described as "a personal collection of knowledge", which helps us to make efficient decisions

Another explanation[1]



Data represents a fact or statement of event without relation to other things.

Information embodies the understanding of a relationship of some sort, possibly cause and effect.

Knowledge represents a pattern that connects and generally provides a high level of predictability as to what is described or what will happen next.

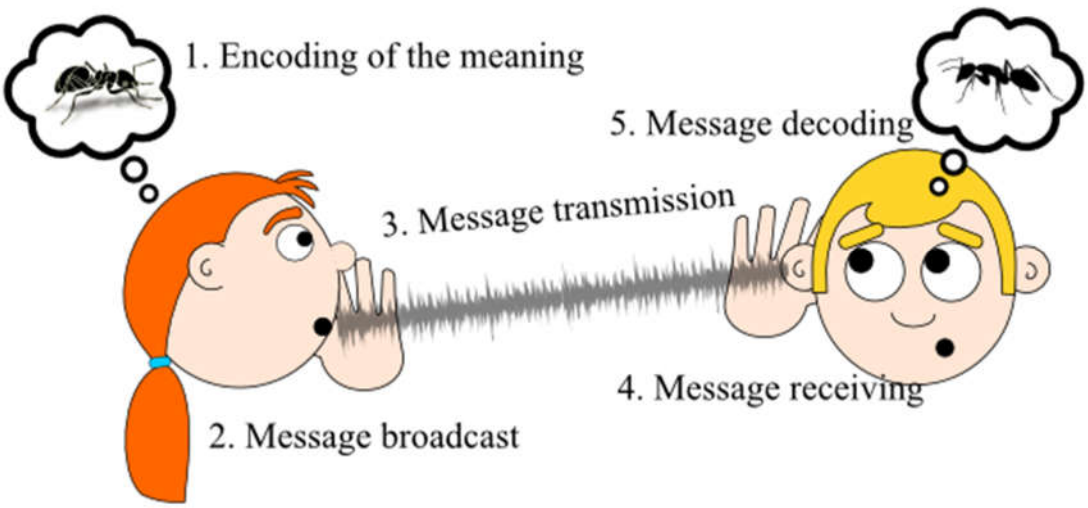
Example

Data	34000, 85332, 9233, 75544, 8234, 9120, 72000
Context	quarter sales (in thousands of Euro) for the products A, B, C, D, E, F, G
Information	quarter sales of A are 34000 €, quarter sales of B are 85332 € …
Knowledge	the best selling product is B, while the worst one is E - to make more money we should focus on B in the production, while E should be eliminated from it
Wisdom	we should collect data about sales to find out, which product is the best selling, and prefer it in the production, while also conducting a research to identify the key reasons why the product is more successful; then, based on that we should consider these reasons in the new series of products A2, B2, C2 …

[1] http://www.systems-thinking.org/dikw/dikw.htm

Communication

- meaningful exchange of information between two or more participants
- for exchange are used various forms of signals: electromagnetic waves, chemical interactions, mechanical interaction …
- Phases:
   0. **Thought:** Information exists in the mind of the sender. This can be a concept, idea, information, or feeling.
   1. **Encoding**: The information is encoded into a message using words or other symbols.
   2. **Broadcast**: The message is turned into a signal = physical phenomenon – sound, electromagnetic waves, mechanical motion …
   3. **Transmission**: The physical representation of the message is transmitted over the medium (e.g. speech – sound through the air). The representation is exposed to noise = external modifications of the original signal.
   4. **Receiving**: The receiver intercepts the signal and turns it into a message.
   5. **Decoding**: Lastly, the receiver translates the message – i.e. words or symbols – into a concept or information that a person can understand.



Units of Data

- **Unit of data** = a measurable portion of data.
- **Hartley-Shannon principle of storage**: the information that can be stored in a system with N different states is proportional to $\log_{base}N$. The choice of the base determines the units
  - if base = 2, then the unit is a **bit**;
  - if base = 3, then the unit is a **trit** (one trit = $\log_2(3)$ bits)**;**
  - if base = 10, then the unit is a **dit** (one trit = $\log_2(10)$ bits)**;**
  - If base = 16, then the unit is a **hexit** (one hexit = $\log_2(16)$ bits)**.**
- The simplest system – 2 states – bits are the most frequent unit.
- Representation of bits

Principle	0	1
Electric Switch	off	on
Voltage	< 0.4V	>2.6 V
Reflection of light (CD, DVD)	pits (poor reflection)	lands (high reflection)

Qubit

- Essential unit of quantum information
- It can be 0, 1, or a superposition ("combination") of both

Standard data units

- **Bit –** the smallest unit of data. Single bit can hold only one of two values: either *zero* or *on*, which are represented by 2 distinct physical states (switch is on/off; 2 different voltage levels – integrated circuits; two directions of magnetization – hard disk, magnetic tape; 2 distinct levels of light intensity – optical discs …). Its symbol is the lower case **b**.
- **Byte –** a larger unit – nowadays a combination of 8 bits (octet); in the past there were used also different lengths (4, 6, 7, 9). It is also a term, which names a data type in programming languages with such as length.
- There are also units, which are adopted mainly by software developers
  - ***Nibble*** = 4 b – a half a byte/octet. In the hexadecimal notation it is represented by one hexadecimal digit (0 – F).
  - ***Word*** = The *x86* platform (Intel and AMD CPUs) originally had a word size of 16 bits (2 bytes) and that usage of the term is confusingly retained even though the actual processor word size is now 64 bits. Therefore a *dword (double word)* designates a 32-bit (4-byte) unit, and a *qword (quadruple word)* means a 64 bit long unit[2].

[2] http://en.wikipedia.org/wiki/Word_%28computing%29

Byte is still too small for a number of situations; therefore there are larger units derived from the byte using the prefixes like kilo, mega etc.

However, there is confusion in their meaning:

- producers of computer storage devices prefer the original SI values (**powers of 10**);
- software and computer industries often use binary estimates of the SI-prefixed quantities (so the kilo- means in such as case 1024).

The confusion causes a lot of problems.

Example: difference between the capacity declared by the producer and its value reported by the operating system. A 1 TB hard disk drive would be reported as $\frac{10^{12}}{2^{30}}$ *= 931 GB drive.*

There was made an attempt to make the difference clear, so the binary units with binary prefixes were defined.

SI Prefixes	
1 kB	$= 10^3$ B = 1000 B
1 MB	$= 10^6$ B = 1000 kB
1 GB	$= 10^9$ B = 1000 MB
1 TB	$= 10^{12}$ B = 1000 GB
…	

Binary Prefixes (IEC standard)		
1 KiB	=	2^{10} B = 1024 B
1 MiB	=	2^{20} B = 1024 KiB
1 GiB	=	2^{30} B = 1024 MiB
1 TiB	=	2^{40} B = 1024 GiB
…		

Exercises

1. Convert the units
   a. 14 kB = ? b
   b. 84 MiB = ? B
   c. 1024 b = ? B
   d. 32 000 b = ? KiB
   e. 96 GB = ? b
   f. 57 MB = ? B
   g. 1 048 576 b = MiB
   h. 65536 MiB = ? GiB

2. Let's assume an Internet connection at 24 Mbps (megabits per second). We have 10 minutes to download our photos from the Google Drive. How much data can be downloaded within the given time?

3. My photo is stored in a picture, which has 1600 x 1200 pixels (i.e. tiny points the picture is made of). Each pixel is stored in 24 bits? How many **MiB** does it take?