Introduction to Computer Science

By:
Huong Nguyen

Introduction to Computer Science

By:
Huong Nguyen

Online:
< http://cnx.org/content/col10776/1.1/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: July 29, 2009
PDF generated: October 27, 2012
For copyright and attribution information for the modules contained in this collection, see p. 157.

Table of Contents

Chapter 1

Introduction to Computer Science

1.1 Basic concepts[1]

1.1.1 Information & Information Processing

1.1.1.1 Data – Information – Knowledge

The content of the human mind can be classified into four categories:

- Data: symbols;
- Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions;
- Knowledge: understanding of data and information; answers "how" questions;
- Wisdom: evaluated understanding.

Data

Data consist of raw facts and figures - it does not have any meaning until it is processed and turned into something useful.

Data comes in many forms; the main ones are letters, numbers and symbols.

Data is a prerequisite to information. For example, the two data items below could represent some very important information:

```
    DATA                INFORMATION
123424331911       Your winning Lottery ticket number
211192               Your Birthday
```

An organization sometimes has to decide on the nature and volume of data that is required for creating the necessary information.

Information

Information is the data that has been processed in such a way as to be meaningful to the person who receives it.

```
INFORMATION = DATA + CONTEXT + MEANING
```

Example

Consider the number19051890. It has no meaning or context. It is an instance of data.

If a context is given : it is a date (Vietnamese use French format ddmmyyyy). This allows us to register it as 19th May 1890. It still has no meaning and is therefore not information

Meaning : The birth date of Vietnamese President Ho Chi Minh.

[1]This content is available online at <http://cnx.org/content/m27715/1.1/>.

This gives us all the elements required for it to be called 'information'

Knowledge

By knowledge we mean the human understanding of a subject matter that has been acquired through proper study and experience.

Knowledge is usually based on learning, thinking, and proper understanding of the problem area. It can be considered as the integration of human perceptive processes that helps them to draw meaningful conclusions.

Consider this scenario: A person puts his finger into very hot water.

Data gathered: The finger nerve sends pain data to the brain.

Processing: The brain considers the data and comes up with...

Information: The painful finger means it is not in a good place.

Action: The brain tells finger to remove itself from hot water.

Knowledge: Sticking the finger in hot water is a bad idea.

Knowledge is having an understanding of the "rules".

The terms Data, Information, Knowledge, and Wisdom are sometimes presented in a form that suggests a scale.

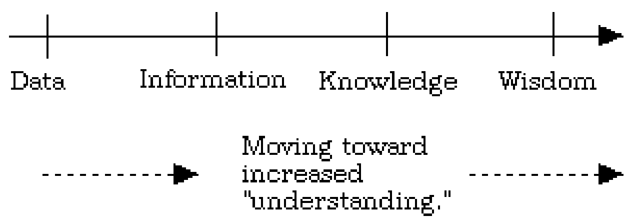

Figure 1.1: Data, Information, knowledge, wisdom along a scale

1.1.1.2 Information Processing

Information processing is the change (processing) of information in any manner detectable by an observer. Information processing may more specifically be defined in terms by Claude E. Shannon as the conversion of latent information into manifest.Input, process, output is a typical model for information processing. Each stage possibly requires data storage.

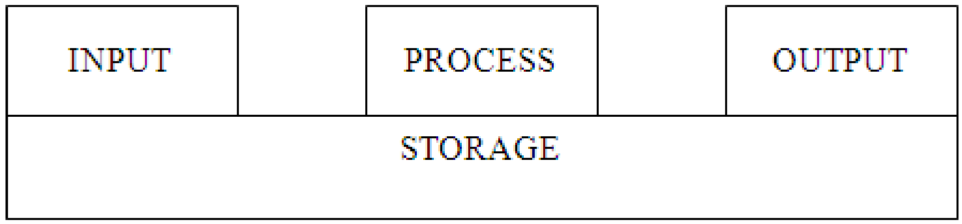

Figure 1.2: Model of information processing

Now that computer systems have become so powerful, some have been designed to make use of information in a knowledgeable way. The following definition is of information processing

The electronic capture, collection, storage, manipulation, transmission, retrieval, and presentation of information in the form of data, text, voice, or image and includes telecommunications and office automation functions.

History and Classification of Computers

1.1.1.3 History of Computers

Webster's Dictionary defines "computer" as any programmable electronic device that can store, retrieve, and process data.

Blaise Pascal invents the first commercial calculator, a hand powered adding machine

In 1946, ENIAC, based on John Von Neuman model completes.The first commercially successful computer is IBM 701.

A generation refers to the state of improvement in the development of a product. This term is also used in the different advancements of computer technology. With each generation, the circuitry has gotten smaller and more advanced than the previous generations before it. As a result of the miniaturization, the speed, power and memory of computers has proportionally increased. New discoveries are constantly being developed that affect the way we live, work and play. In terms of technological developments over time, computers have been broadly classed into five generations.

1.1.1.3.1 The First Generation - 1940-1956

The first computers used vacuum tubes for circuitry and magnetic drums for memory, and were often enormous, taking up entire rooms. They were very expensive to operate and in addition to using a great deal of electricity, they generated a lot of heat, which was often the cause of malfunctions. First generation computers relied on machine language to perform operations, and they could only solve one problem at a time. Input was based on punched cards and paper tape, and output was displayed on printouts.

The computers UNIVAC , ENIAC of the US and BESEM of the former Soviet Union are examples of first-generation computing devices.

1.1.1.3.2 The Second Generation - 1956-1963

Transistors replaced vacuum tubes and ushered in the second generation of computers. Computers becomed smaller, faster, cheaper, more energy-efficient and more reliable than their first-generation predecessors.

Second-generation computers still relied on punched cards for input and printouts for output. High-level programming languages were being developed, such as early versions of COBOL and FORTRAN.

The first computers of this generation were developed for the atomic energy industry.

The computers IBM-1070 of the US and MINSK of the former Soviet Union belonged to the second generation.

1.1.1.3.3 The Third Generation - 1964-1971: Integrated Circuits

The development of the integrated circuit was the hallmark of the third generation of computers. Transistors were miniaturized and placed on silicon chips, called semiconductors, which drastically increased the speed and efficiency of computers. Users interacted with third generation computers through keyboards and monitors and interfaced with an operating system, which allowed the device to run many different applications at one time. Typical computers of the third generation are IBM 360 (United States) and EC (former Soviet Union).

1.1.1.3.4 The Fourth Generation - 1971-Present: Microprocessors

The microprocessor brought the fourth generation of computers, as thousands of integrated circuits were built onto a single silicon chip. What in the first generation filled an entire room could now fit in the palm of the hand. The Intel 4004 chip, developed in 1971, located all the components of the computer - from the central processing unit and memory to input/output controls - on a single chip.

In 1981 IBM introduced its first computer for the home user, and in 1984 Apple introduced the Macintosh. Microprocessors also moved out of the realm of desktop computers and into many areas of life as more and more everyday products began to use microprocessors.

As these small computers became more powerful, they could be linked together to form networks, which eventually led to the development of the Internet. Fourth generation computers also saw the development of GUI (Graphic User Interface), the mouse and handheld devices.

1.1.1.3.5 The Fifth Generation - Present and Beyond: Artificial Intelligence

Fifth generation computing devices, based on artificial intelligence, are still in development, though there are some applications, such as voice recognition, that are being used today. The use of parallel processing and superconductors is helping to make artificial intelligence a reality. Quantum computation and molecular and nanotechnology will radically change the face of computers in years to come. The goal of fifth-generation computing is to develop devices that respond to natural language input and are capable of learning and self-organization.

1.1.1.4 Classification of Computers

Computers are available in different shapes, sizes and weights, due to these different shapes and sizes they perform different sorts of jobs from one another.

- **Mainframe and Super Computers**

The biggest in size, the most expensive in price than any other is classified and known as super computer. It can process trillions of instructions in seconds. Governments specially use this type of computer for their different calculations and heavy jobs. This kind of computer is also helpful for forecasting weather reports worldwide.

Another giant in computers after the super computer is Mainframe, which can also process millions of instruction per second and capable of accessing billions of data. This computer is commonly used in big hospitals, airline reservations companies, and many other huge companies prefer mainframe because of its capability of retrieving data on a huge basis. This is normally too expensive and out of reach from a salary-based person who wants a computer for his home.

- **Minicomputers**

This computer offers less than mainframe in work and performance. These are the computers, which are mostly preferred by the small type of business personals, colleges, and so on.

- **Microcomputers**

These computers are lesser in cost than the computers given above and also, small in size; They can store a big amount of data and have a memory to meet the assignments of students and other necessary tasks of business people. There are many types of microcomputers: desktop, workstation, laptop, PDA , etc.

1.1.1.5 Computer Science and Relevant Sciences

In 1957 the German computer scientist Karl Steinbuch coined the word informatik by publishing a paper called Informatik: Automatische Informationsverarbeitung (i.e. "Informatics: automatic information processing"). The French term informatique was coined in 1962 by Philippe Dreyfus together with various translations—informatics (English), informatica (Italian, Spanish, Portuguese), informatika (Russian) referring to the application of computers to store and process information.

The term was coined as a combination of "information" and "automation", to describe the science of automatic information processing.

Informatics is more oriented towards mathematics than computer science.

1.1.1.5.1 Computer Science

Computer Science is the study of computers, including both hardware and software design. Computer science is composed of many broad disciplines, for instance, artificial intelligence and software engineering.

1.1.1.5.2 Information Technology

Includes all matters concerned with the furtherance of computer science and technology and with the design, development, installation, and implementation of information systems and applications

1.1.1.5.3 Information and Communication Technology

ICT (information and communications technology - or technologies) is an umbrella term that includes any communication device or application, encompassing: radio, television, cellular phones, computer and network hardware and software, satellite systems and so on, as well as the various services and applications associated with them, such as videoconferencing and distance learning.

1.2 Data Representation in a Computer[2]

Computer must not only be able to carry out computations, they must be able to do them quickly and efficiently. There are several data representations, typically for integers, real numbers, characters, and logical values.

1.2.1 Number Representation in Various Numeral Systems

A numeral system is a collection of symbols used to represent small numbers, together with a system of rules for representing larger numbers. Each numeral system uses a set of digits. The number of various unique digits, including zero, that a numeral system uses to represent numbers is called base or radix.

[2]This content is available online at <http://cnx.org/content/m27737/1.1/>.

1.2.1.1 Base - b numeral system

b basic symbols (or digits) corresponding to natural numbers between 0 and b − 1 are used in the representation of numbers.

To generate the rest of the numerals, the position of the symbol in the figure is used. The symbol in the last position has its own value, and as it moves to the left its value is multiplied by b.

We write a number in the numeral system of base b by expressing it in the form

$$N_{(b)} = a_n a_{n-1} a_{n-2} ... a_1 a_0 a_{-1} a_{-2} ... a_{-m} \tag{1.1}$$

N(b), with n+1 digit for integer and m digits for fractional part, represents the sum:

$$N_{(b)} = a_n.b^n + a_{n-1}.b^{n-1} + a_{n-2}.b^{n-2} + ... + a_1.b^1 + a_0.b^0 + a_{-1}.b^{-1} + a_{-2}.b^{-2} + ... + a_{-m}.b^{-m}$$

or

$$N_{(b)} = \sum_{i=-m}^{n} a_i.b^i$$

Figure 1.3

in the decimal system. Note that a_i is the i^{th} digit from the position of a_0

Decimal, Binary, Octal and Hexadecimal are common used numeral system. The decimal system has ten as its base. It is the most widely used numeral system, because humans have four fingers and a thumb on each hand, giving total of ten digit over both hand.

Switches, mimicked by their electronic successors built of vacuum tubes, have only two possible states: "open" and "closed". Substituting open=1 and closed=0 yields the entire set of binary digits. Modern computers use transistors that represent two states with either high or low voltages. Binary digits are arranged in groups to aid in processing, and to make the binary numbers shorter and more manageable for humans.Thus base 16 (hexadecimal) is commonly used as shorthand. Base 8 (octal) has also been used for this purpose.

Decimal System

Decimal notation is the writing of numbers in the base-ten numeral system, which uses various symbols (called digits) for no more than ten distinct values (0, 1, 2, 3, 4, 5, 6, 7, 8 and 9) to represent any number, no matter how large. These digits are often used with a decimal separator which indicates the start of a fractional part, and with one of the sign symbols + (positive) or − (negative) in front of the numerals to indicate sign.

Decimal system is a place-value system. This means that the place or location where you put a numeral determines its corresponding numerical value. A two in the one's place means two times one or two. A two in the one-thousand's place means two times one thousand or two thousand.

The place values increase from right to left. The first place just before the decimal point is the one's place, the second place or next place to the left is the ten's place, the third place is the hundred's place, and so on.

The place-value of the place immediately to the left of the "decimal" point is one in all place-value number systems. The place-value of any place to the left of the one's place is a whole number computed from a product (multiplication) in which the base of the number system is repeated as a factor one less number of times than the position of the place.

For example, 5246 can be expressed like in the following expressions

$$\begin{aligned} 5246 &= 5 \times 10^3 + 2 \times 10^2 + 4 \times 10^1 + 6 \times 10^0 \\ &= 5 \times 1000 + 2 \times 100 + 4 \times 10 + 6 \times 1 \end{aligned} \tag{1.2}$$

The place-value of any place to the right of the decimal point is a fraction computed from a product in which the reciprocal of the base—or a fraction with one in the numerator and the base in the denominator—is repeated as a factor exactly as many times as the place is to the right of the decimal point.

For example

$$\begin{aligned} 254.68 &= 2 \times 10^2 + 5 \times 10^1 + 4 \times 10^0 + 6 \times 10^{-1} + 8 \times 10^{-2} \\ &= 200 + 50 + 4 + \tfrac{6}{10} + \tfrac{8}{100} \end{aligned} \tag{1.3}$$

Image not finished

1.2.1.2 Binary System

The binary number system is base 2 and therefore requires only two digits, 0 and 1. The binary system is useful for computer programmers, because it can be used to represent the digital on/off method in which computer chips and memory work.

A binary number can be represented by any sequence of bits (binary digits), which in turn may be represented by any mechanism capable of being in two mutually exclusive states.

Counting in binary is similar to counting in any other number system. Beginning with a single digit, counting proceeds through each symbol, in increasing order. Decimal counting uses the symbols 0 through 9, while binary only uses the symbols 0 and 1.

When the symbols for the first digit are exhausted, the next-higher digit (to the left) is incremented, and counting starts over at 0A single bit can represent one of two values, 0 or 1.Binary numbers are convertible to decimal numbers.

Here's an example of a binary number, $11101.11_{(2)}$, and its representation in the decimal notation

Position of the separator character

Binary Number	1	1	1	0	1	.	1	1
Position	4	3	2	1	0		-1	-2
Place Value	2^4	2^3	2^2	2^1	2^0		2^{-1}	2^{-2}
Decimal Number	16	8	4	2	1		0.5	0.25

So,

$$\mathbf{11101.11_{(2)} = 1x16 + 1x8 + 1x4 + 0x2 + 1x1 + 1x0.5 + 1x0.25 = 29.75_{(10)}}$$

Another Conversion

$$\mathbf{10101_{(2)} = 1x2^4 + 0x2^3 + 1x2^2 + 0x2^1 + 1x2^0 = 16 + 0 + 4 + 0 + 1 = 21_{(10)}}$$

Figure 1.4

$$235.64_{(8)} = 2 \times 8^2 + 3 \times 8^1 + 5 \times 8^0 + 6 \times 8^{-1} + 4 \times 8^{-2} = 157.8125_{(10)} \tag{1.4}$$

1.2.1.3 Hexadecimal System

The hexadecimal system is base 16. Therefore, it requires 16 digits. The digits 0 through 9 are used, along with the letters A through F, which represent the decimal values 10 through 15. Here is an example of a hexadecimal number and its decimal equivalent:

$$34F5C_{(16)} = 3 \times 16^4 + 4 \times 16^3 + 15 \times 16^2 + 5 \times 16^1 + 12 \times 16^0 = 216294_{(10)} \tag{1.5}$$

The hexadecimal system (often called the hex system) is useful in computer work because it is based on powers of 2. Each digit in the hex system is equivalent to a four-digit binary number. Table below shows some hex/decimal/binary equivalents.

Hexadecimal Digit	Decimal Equivalent	Binary Equivalent
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111
8	8	1000
9	9	1001
A	10	1010
B	11	1011
C	12	1100
D	13	1101
E	14	1110
F	15	1111
10	16	10000
F0	240	11110000
FF	255	11111111

Table 1.1

1.2.1.4 Octal System

Binary is also easily converted to the octal numeral system, since octal uses a radix of 8, which is a power of two (namely, 23, so it takes exactly three binary digits to represent an octal digit). The correspondence between octal and binary numerals is the same as for the first eight digits of hexadecimal in the table above. Binary 000 is equivalent to the octal digit 0, binary 111 is equivalent to octal 7, and so forth.

Converting from octal to binary proceeds in the same fashion as it does for hexadecimal:

$$65_{(8)} = 110\,1012 \tag{1.6}$$

$$17_{(8)} = 001\ 1112 \tag{1.7}$$

And from octal to decimal:

$$235.64_{(8)} = 2 \times 8^2 + 3 \times 8^1 + 5 \times 8^0 + 6 \times 8^{-1} + 4 \times 8^{-2} = 157.8125_{(10)} \tag{1.8}$$

1.2.1.5 Converting from decimal to base–b

To convert a decimal fraction to another base, say base b, you split it into an integer and a fractional part. Then divide the integer by b repeatedly to get each digit as a remainder. Namely, with value of integer part $= d_{n-1}d_{n-2}...d_2d_1d_{0(10)}$, first divide value by b the remainder is the least significant digit a_0 . Divide the result by b, the remainder is a_1 .Continue this process until the result is zero, giving the most significant digit, a_{n-1} . Let's convert $43868_{(10)}$ to hexadecimal:

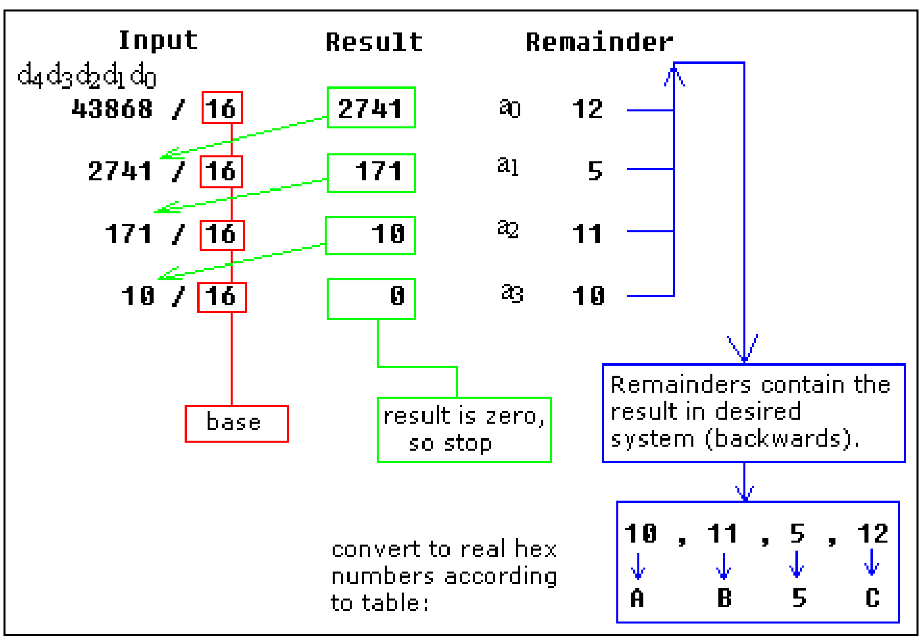

Figure 1.5: Converting from decimal to hexadecimal

After that, multiply the fractional part by b repeatedly to get each digit as an integer part. We will continue this process until we get a zero as our fractional part or until we recognize an infinite repeating pattern.

Now convert 0.625 to hexadecimal :
.
0.39625 * 16 = 0.625 ——————————————-> 0
.625* 16 = 10 ——————————> A.
We get fractional part is zero.
In summary, the result of conversion $43868.39625_{(10)}$ to hexadecimal is AB5C.0A

1.2.2 Data Representation in a Computer. Units of Information

1.2.2.1 Basic Principles

Data Representation refers to the methods used internally to represent information stored in a computer. Computers store lots of different types of information:

- numbers
- text
- graphics of many varieties (stills, video, animation)
- sound

At least, these all seem different to us. However, all types of information stored in a computer are stored internally in the same simple format: a sequence of 0's and 1's. How can a sequence of 0's and 1's represent things as diverse as your photograph, your favorite song, a recent movie, and your term paper?

- Numbers must be expressed in binary form following some specific standard.
- Character data are assigned a sequence of binary digits
- Other types of data, such as sounds, videos or other physical signals are converted to digital following the schema below

Digital signal
Continuous signalPhysical signalComputerConvert ADSensor

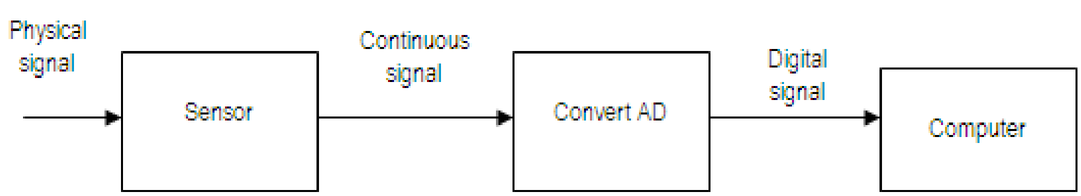

Figure 1.6: Process of converting from physical signal to digital signal

Depending on the nature of its internal representation, data items are divided into:

- Basic types (simple types or type primitives) : the standard scalar predefined types that one would expect to find ready for immediate use in any programming language
- Structured types(Higher level types) are then made up from such basic types or other existing higher level types.

1.2.2.2 Units of Information

The most basic unit of information in a digital computer is called a BIT, which is a contraction of Binary Digit. In the concrete sense, a bit is nothing more than a state of "on" or "off" (or "high" and "low") within a computer circuit. In 1964, the designers of the IBM System/360 mainframe computer established a convention of using groups of 8 bits as the basic unit of addressable computer storage. They called this collection of 8 bits a byte.

Computer words consist of two or more adjacent bytes that are sometimes addressed and almost always are manipulated collectively. The word size represents the data size that is handled most efficiently by a particular architecture. Words can be 16 bits, 32 bits, 64 bits, or any other size that makes sense within the context of a computer's organization.

Some other units of information are described in the following table :

Name	Symbol	Value	Base 16	Base 10
kilo	k/K	2^{10} = 1,024	= $16^{2.5}$	> 10^3
mega	M	2^{20} = 1,048,576	= 16^5	> 10^6
giga	G	2^{30} = 1,073,741,824	= $16^{7.5}$	> 10^9
tera	T	2^{40} = 1,099,511,627,776	= 16^{10}	> 10^{12}
peta	P	2^{50} = 1,125,899,906,842,624	= $16^{12.5}$	> 10^{15}
exa	E	2^{60} = 1,152,921,504,606,846,976	= 16^{15}	> 10^{18}
zetta	Z	2^{70} = 1,180,591,620,717,411,303,424	= $16^{17.5}$	> 10^{21}
yotta	Y	2^{80} = 1,208,925,819,614,629,174,706,176	= 16^{20}	> 10^{24}

Figure 1.7

Representation of Integers

An integer is a number with no fractional part; it can be positive, negative or zero. In ordinary usage, one uses a minus sign to designate a negative integer. However, a computer can only store information in bits, which can only have the values zero or one. We might expect, therefore, that the storage of negative integers in a computer might require some special technique - allocating one sign bit (often the most significant bit) to represent the sign: set that bit to 0 for a positive number, and set to 1 for a negative number.

1.2.2.3 Unsigned Integers

Unsigned integers are represented by a fixed number of bits (typically 8, 16, 32, and/or 64)

- With 8 bits, 0…255 (0016…FF16) can be represented;
- With 16 bits, 0…65535 (000016…FFFF16) can be represented;
- In general, an unsigned integer containing n bits can have a value between 0 and $2^n - 1$

If an operation on bytes has a result outside this range, it will cause an 'overflow'