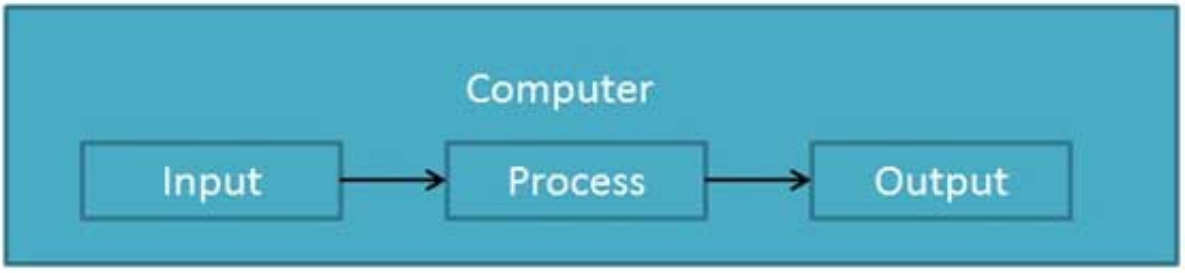
Linux Programming for Bioinformatics

Contents

Computer Science - Overview

We are living in an information-rich world and it is becoming a necessity to know something about the computer science, especially for a student majored in CS-related field, such as bioinformatics. Here we will introduce some fundamental knowledge about the computer science and related.

Functionalities of a computer



Any digital computer can carry out 5 functions in gross terms:

```
- Takes data as input (输入).
- Stores the data/instructions in its memory and use them when requi
red (存储).
- Processes the data and converts it into useful information (处理).
- Generates the output (输出)
- Controls all the above four steps (控制).
```

In a word, a computer is an electronic data processing device which

```
- accepts and stores data input,
- processes the data input, and
- generates the output in required formats.
```

Computer - history

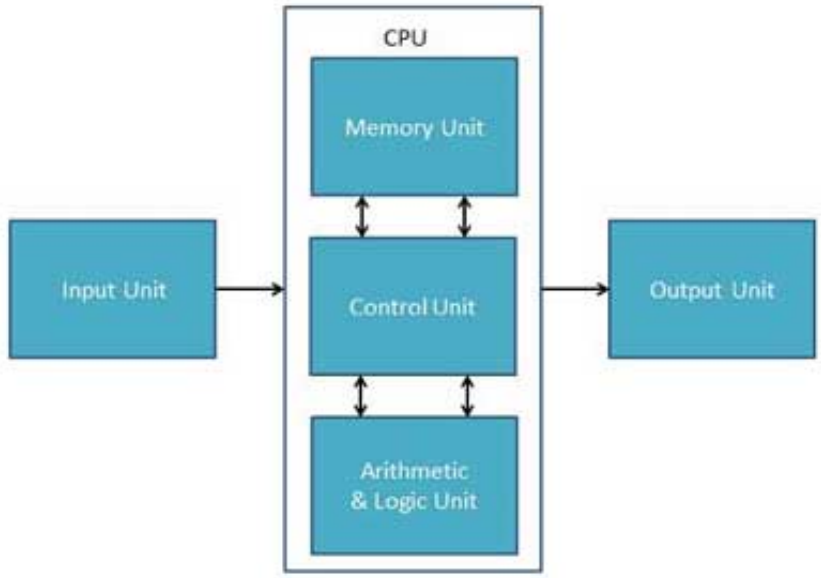
S.N.	Generations & Description
1	**First Generation**: 1946-1959. Vacuum tube (真空管) based.
2	**Second Generation**: 1959-1965. Transistor (晶体管) based.
3	**Third Generation**: 1965-1971. Integrated Circuit (集成电路) based.
4	**Fourth Generation**: 1971-1980. VLSI microprocessor (大规模集成电路微处理器) based.
5	**Fifth Generation**: 1980-onwards. ULSI microprocessor (超大规模集成电路微处理器) based

Computer - types

Computers can be broadly classified by their speed and computing power:

Type	Specifications
PC (个人计算机)	Single-user computer system having moderately powerful microprocessor
WorkStation (工作站)	Single-user computer system which is similar to PC but have more powerful microprocessor.
Mini Computer (小型机)	Multi-user computer system capable of supporting hundreds of users simultaneously.
Main Frame (大型机)	Multi-user computer system supporting hundreds of users simultaneously. Software technology is different from minicomputer.
Supercomputer (超级计算机)	Extremely fast computer which can execute hundreds of millions of instructions per second.

Computer - components



Input unit

This unit contains devices with the help of which we enter data into computer. This unit makes link between user and computer. The input devices translate the information into the form understandable by computer.

The important input devices include keyboard, joystick, mouse, scanner, microphone and etc.

Central processing unit (CPU)

CPU is considered as the brain of the computer. CPU performs all types of data processing operations. It stores data, intermediate results and instructions(program). It controls the operation of all parts of computer.

CPU itself has following three components

```
- ALU (Arithmetic Logic Unit)
- Memory or Storage Unit
- Control Unit
```

Output unit

Output unit consists of devices with the help of which we get the information from computer. This unit is a link between computer and users. Output devices translate the computer's output into the form understandable by users.

The important output devices include monitor, printer and graphic plotters.

Computer - Memory

The computer memory is used to store data and instructions. The memory is divided into large number of small parts called cells. Each cell has a unique address which varies from 0 to $MEMSIZE - 1$. For example if computer has 64k words, then this memory unit has 64 * 1024=65536 memory cells. The address of these locations varies from 0 to 65535.

Memory is primarily of three types

- **Cache Memory (高速缓存)**： Cache memory is a very high speed semiconductor memory which can speed up CPU. It acts as a buffer between the CPU and main memory. It is used to hold those parts of data and program which are most frequently used by CPU. The parts of data and programs are transferred from disk to cache memory by operating system, from where CPU can access them.
- **Primary Memory/Main Memory (内存/主存)**：Primary memory holds only those data and instructions on which computer is currently working. It has limited capacity and data is lost when power is switched off. It is generally made up of semiconductor device. These memories are not as fast as registers. The data and instruction required to be processed reside in main memory. It is divided into two subcategories RAM (Random access memory, 随机存取存储) and ROM (只读存储).
- **Secondary Memory (外存)**：This type of memory is also known as external memory or non-volatile. It is slower than main memory. These are used for storing data/Information permanently. CPU directly does not access these memories instead they are accessed via input-output routines. Contents of secondary memories are first transferred to main memory, and then CPU can access it. For example : disk, CD-ROM, DVD etc.

Computer - Memory Units (存储单位)

Memory unit is the amount of data which can be stored in the storage unit and which can be expressed in terms of Bytes.

Unit	Description
Bit (位)	A binary digit is logical 0 and 1 representing a passive or an active state.
Byte (字节)	A group of 8 bits is called byte. A byte is the smallest unit which can represent a data item or a character.
Word (字)	A computer word, like a byte, is a group of fixed number of bits processed as a unit which varies from computer to computer but is fixed for each computer.

The length of a computer word is called word-size or word length and it may be as small as 8 bits or may be as long as 64 bits. A computer stores the information in the form of computer words.

A few higher storage units are following:

Sr.No.	Unit	Description
1	Kilobyte (KB)	1 KB = 2^10 Bytes
2	Megabyte (MB)	1 MB = 2^10 KB
3	GigaByte (GB)	1 GB = 2^10 MB
4	TeraByte (TB)	1 TB = 2^10 GB
5	PetaByte (PB)	1 PB = 2^10 TB

Computer - Hardware

Hardware represents the physical and tangible components of a computer i.e. the components that can be seen and touched.

Examples of Hardware are following:

- **Input devices** -- keyboard, mouse etc.
- **Output devices** -- printer, monitor etc.
- **Secondary storage devices** -- Hard disk, CD, DVD etc.
- **Internal components** -- CPU, motherboard, RAM etc.

Relationship between Hardware and Software

Hardware and software are mutually dependent on each other. Both of them must work together to make a computer produce a useful output.

Software cannot be utilized without supporting hardware.

- Hardware without set of programs to operate upon cannot be utilized and is useless.
- To get a particular job done on the computer, relevant software should be loaded into the hardware
- Hardware is a one-time expense.
- Software development is very expensive and is a continuing expense.
- Different software applications can be loaded on a hardware to run different jobs.
- A software acts as an interface between the user and the hardware.
- If hardware is the 'heart' of a computer system, then software is its 'soul'. Both are complimentary to each other.

Binary Number System

basic concept behing the binary system

In decimal system, the number, 192, can be orgranized as

```
10^2 | 10^1 | 10^0
   1 |    9 |    2
```

Similarly, in binary system

```
2^2 | 2^1 | 2^0
  1 |   0 |   1
```

can be used to represent the decimal number $1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 5$.

Exercise

- (1) What would the binary number 1011 be in decimal notation?
- (2) What would the binary number 11001110 be in decimal notation?

Binary addition

Always remember:

- 0 + 0 = 0
- 0 + 1 = 1 + 0 = 1
- 1 + 1 = 10

Exercise

Using this principle, compute the following binary addition

- (1) 1001 + 1100
- (2) 101 + 1000

Converting decimal to binary number

Let's formalize this algorithm:

```
1.  Let D= the number we wish to convert from decimal to binary.
2.  Repeat until D=0:
    a) If D is odd, put "1" in the leftmost open column, and subtrac
t 1 from D.
    b) If D is even, put "0" in the leftmost open column.
    c) Divide D by 2.
    End Repeat
```

For the decimal number 19, apply the algorithm:

```
1. Let D = 19
2. (a) D is odd: put "1" to "2^0" bit, and D = 19 - 1 = 18
   (c) Divide D by 2: D = 18/2 = 9
   D is not 0, repeat
2. (a) D is odd: put "1" to "2^1" bit, and D = 9 - 1 = 8
   (c) Divide D by 2: D = 8 / 2 = 4
   D is not 0, repeat
2. (b) D is even: put "0" to "2^2" bit
   (c) Divide D by 2: D = 4 / 2 = 2
   D is not 0, repeat
2. (b) D is even: put "0" to "2^3" bit
   (c) Divide D by 2: D = 2/2 = 1
   D is not 0, repeat
2. (a) D is odd: put "1" to "2^4" bit, and D = 1 - 1 = 0
   D is now 0, end repeat

   Now, the decimal number, 19, is converted into 10011.
```

Exercise

- (1) Convert the decimal numbers 27, 95, 167, 323 into binary numbers.

More about octal and hexadecimal system

As we know $2^3 = 8$, and $2^4 = 16$, octal number and hexadecimal number can be easily converted into binary number.

For octal number:

```
000 | 001 | 010 | 011 | 100 | 101 | 110 | 111
  0 |   1 |   2 |   3 |   4 |   5 |   6 |   7
```

While for hexadecimal number:

```
0000|0001|0010|0011|0100|0101|0110|0111|1000|1001|1010|1011|1100|110
1|1110|1111
   0|   1|   2|   3|   4|   5|   6|   7|   8|   9|   A|   B|   C|
D|   E|   F
```

Therefore, each octal digit can be converted into 3-digit binary number; in a similar fashion, every hexadecimal digit can be converted into 4-digit binary number.

Here is an example for octal number, 0372, can be converted to binary number, 11111010:

```
  3 |   7 |   2
011 | 111 | 010
```

Also, the hexadecimal number, 0x36EF, can be converted into binary number, 11011011101111, in this way:

```
   3|    6|    E|    F
0011| 0110| 1110| 1111
```

Note that in computer science, a octal number starts with '0', while a hexadecimal number starts with '0x'.

Using the similar approach, we can easily convert binary numbers to either octal or hexadecimal numbers.

Exercise

- (1) Convert the octal numbers 073245, 0174 into binary and decimal numbers;
- (2) Convert the hexademical numbers 0x34EFA, 0xD5A7, 0xBBCA to binary and decimal numbers