Kiến thức Cơ sở

Môn học: Nhập môn Công nghệ thông tin I

Nội dung

- Tổng quan thông tin
- Hệ thống đếm
- Biểu diễn thông tin trong máy tính
- Hệ thống mã hoá
- Hệ thống tập tin (theo góc nhìn lập trình)

Tổng quan thông tin

Tổng quan thông tin

- Khái niệm
  - Thông tin là tất cả những gì đem lại sự hiểu biết về thế giới xung quanh (sự vật, hiện tượng ….) và về chính con người.

Dữ liệu - Thông tin- Tri thức

- Dữ liệu:
  - Là hình thức thể hiện của thông tin trong mục đích lưu trữ và xử lý nhất định.
- Tri thức:
  - Có ý nghĩa khái quát hơn thông tin.
  - Tri thức là mục đích của nhận thức trên cơ sở tiếp nhận thông tin.
  - Quá trình xử lý thông tin chính là quá trình nhận thức để có tri thức.

Hệ thống đếm

Hệ thống đếm

- Khái niệm
- Hệ đếm cơ số 10
- Hệ đếm cơ số bất kì
- Hệ đếm cơ số 2
- Hệ đếm cơ số 16
- Hệ đếm cơ số 8

Hệ thống đếm

- Khái niệm:
  - Hệ thống đếm là tập hợp các kí hiệu và quy tắc để biểu diễn và xác định giá trị các số.
  - Mỗi hệ đếm có 1 số kí tự hữu hạn. Tổng số kí tự của mỗi hệ đếm được gọi là cơ số (base hay radix), kí hiệu là b.
  - Ví dụ:
    - Hệ đếm cơ số 10: 10 kí số: 0 1 2 3 4 5 6 7 8 9
    - 2112 là 1 số trong hệ 10

Hệ thống đếm

- Hệ đếm cơ số 10:
  - Gồm 10 kí số: 0 1 2 3 4 5 6 7 8 9
  - Ví dụ:

$21.12 = 2*10^{1} + 1*10^{0} + 1*10^{-1} + 2*10^{-2}$

$21.12 = 2*10 + 1*1 + 1*1/10 + 2*1/100$

$21.12 = 20 + 1 + 0.1 + 0.02 = 21.12$

Hệ đếm cơ số b bất kì

- Tổng quát:
  - Có b kí tự để thể hiện giá trị số. Kí tự nhỏ nhất là 0, lớn nhất là b-1.
  - Số $N_{(b)}$ trong hệ đếm cơ số b được biểu diễn như sau:

    $N_{(b)} = a_n a_{n-1} \ldots a_0 a_{-1} \ldots a_{-m}$

  và có giá trị:

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

Hệ đếm cơ số b bất kì

– Trong đó:

- b là cơ sở của biểu diễn, $b \in N$, $b \geq 2$.
- a_i là các ký số và $a_i \in N$, $0 \leq i \leq n$, $0 \leq a_i < b$.
- Cách viết trên được gọi là biểu diễn cơ sở b của a.
- Chiều dài của biểu diễn bằng n + 1.
- Nếu có số lẻ thì vị trí đầu tiên sau dấu phẩy là -1, các vị trí tiếp theo là -2, -3, …

Hệ đếm cơ số 2(hệ nhị phân)

- Gồm 2 kí số: 0 1
- Ví dụ:

$$1010.11_2 = 1*2^3+0*2^2+1*2^1+0*2^0+1*2^{-1}+1*2^{-2}$$

$$= 8+0+2+0+0.5+0.25 = 10.75_{10}$$

Hệ đếm cơ số 2(hệ nhị phân)

- Các phép toán:
  - Phép cộng
  - Phép trừ
  - Phép nhân
  - Phép chia

Hệ đếm cơ số 2(hệ nhị phân)

- **Phép cộng**:
  - Cộng có nhớ các cặp số cùng vị trí từ phải sang trái
  - Bảng cộng

+	0	1
0	0	1
1	1	10

Ví dụ:

$$\begin{array}{cccccc} & & 1 & 1 & 1 & 0 \\ + & & 1 & 0 & 0 & 0 \\ \hline 1 & & 0 & 1 & 1 & 0 \end{array}$$

Hệ đếm cơ số 2(hệ nhị phân)

- **Phép trừ:**
  - Số bù 1: đảo tất cả các bit của 1 số nhị phân ta được số bù 1 của nó.
  - Số bù 2: lấy số bù 1 cộng 1 ta được số bù 2 của số nhị phân ban đầu.
  - Ví dụ: x= 1010
  - Số bù 1 của x: 0101
  - Số bù 2 của x: 0110

Hệ đếm cơ số 2(hệ nhị phân)

- **Phép trừ:**
  - Cho 2 số nhị phân x và y, phép trừ: x-y ⇔x+ số bù 2 của y
  - Ví dụ: x=1010, y=0101
  - Số bù 1 của y: 1010
  - Số bù 2 của y: 1011(y_2)
  - X-y=x+y_2=1010+1011=0101

Hệ đếm cơ số 2(hệ nhị phân)

- **Phép nhân:** nhân từ phải qua trái theo cách thông thường
  - Bảng nhân

x	0	1
0	0	0
1	0	1

Ví dụ:

```
      1 0 1 1
x
        1 0
    ---------
      0 0 0 0
    1 0 1 1
    ---------
    1 0 1 1 0
```

Hệ đếm cơ số 2(hệ nhị phân)

- **Phép chia:** trong hệ nhị phân thực hiện tương tư như phép chia trong hệ cơ số 10.
- Ví dụ:

```
 1  1  1  0  1 | 1  0  1
-                1  0  1
 1  0  1
 ---------
 0  1  0  0
   -
    0  0  0
    ---------
    1  0  0  1
      -
       1  0  1
       ---------
       1  0  0 ← Số dư
```

Hệ đếm cơ số 16 (hệ thập lục phân)

- **Gồm 16 kí số:** 0 1 2 3 4 5 6 7 8 9 A B C D E F
- **Ví dụ:**

$3F.2 = 3 \times 16^1 + 15 \times 16^0 + 2 \times 16^{-1}$

$=48 + 15 + 0.125$

$=63.125_{10}$

Hệ đếm cơ số 16 (hệ thập lục phân)

- **Các phép toán:** được thực hiện tương tự như ở hệ thập phân.
- *(Xem tài liệu để tham khảo thêm)*

Hệ đếm cơ số 8 (hệ bát phân)

- **Gồm 8 kí số:** 0 1 2 3 4 5 6
- **Ví dụ:**

$$21.21_8 = 2 \times 8^1 + 1 \times 8^0 + 2 \times 8^{-1} + 1 \times 8^{-2}$$
$$= 16 + 1 + 0.125 + 0.015625$$
$$= 18.140625_{10}$$

Hệ đếm cơ số 8 (hệ bát phân)

- **Các phép toán:** được thực hiện tương tự như ở hệ thập phân.
- *(Xem tài liệu để tham khảo thêm.)*

Chuyển đổi giá trị giữa các hệ đếm

- **Đặc điểm:**
  - Con người sử dụng hệ thập phân.
  - Máy tính sử dụng hệ nhị phân, bát phân và thập lục phân.
- **Nhu cầu:**
  - Chuyển đổi qua lại giữa các hệ đếm.
  - Hệ đếm cơ số 10 sang cơ số 2.
  - Hệ đếm cơ số 16 sang cơ số 10.

Chuyển từ cơ số b sang cơ số 10

- **Cách 1:**
  - Khai triển biểu diễn và tính giá trị biểu thức.
  - Ví dụ:

$1011.01_2 = 1x2^3 + 0x2^2 + 1x2^1 + 1x2^0 + 0x2^{-1} + 1x2^{-2}$

$1011.0 = 8 + 0 + 2 + 1 + 0 + 0.25 = 11.25_{10}$

Chuyển từ cơ số b sang cơ số 10

- **Cách 2:**
  - Nhân chia lồng nhau.
  - Ví dụ:

$\underline{1011}.01_2 = \underline{((1x2 + 0)x2 + 1)x2 + 1} + (1/2 + 0)/2$

$1011.01_2 = \underline{11} + 0.25 = 11.25_{10}$

Chuyển từ cơ số 10 sang cơ số b

- **Đổi phần nguyên:**
  - Chia phần nguyên của cho b và tiếp tục lấy phần nguyên của kết quả chia cho b.
  - Thực hiện cho đến khi thương của phép chia là 0
  - Dãy các số dư ở mỗi lần chia là a_0, a_1, ..., a_n.
  - Phần nguyên của số hệ cở sở b là (a_n...a_1a_0).
- **Đổi phần lẻ:**
  - Nhân phần lẻ cho b và tiếp tục lấy phần lẻ của kết quả nhân cho b. Tiếp tục cho tới khi nào phần lẻ của tích là 0
  - Dãy các số nguyên ở mỗi lần nhân là a_{-1}, a_{-2}, ..., a_{-m} tạo thành phần lẻ ở hệ cơ sở b.

Chuyển từ cơ số 10 sang cơ số b

– **Ví dụ:** đổi 21.125_{10} sang hệ nhị phân (b = 2)

Đổi phần nguyên

21	2			
1	10	2		
	0	5	2	
		1	2	2
			0	1	2
				1	*0*

Đổi phần lẻ

0.125 x 2 = **0**.25

0.25 x 2 = **0**.5

0.5 x 2 = **1**.*0*

Kết quả: 21.125= 10101.001

Chuyển từ cơ số 2 sang cơ số b

- **Từ nhị phân sang thập lục phân:**
  - Nhóm từng bộ 4 bit trong biểu diễn nhị phân rồi chuyển sang kí số tương ứng trong hệ thập lục phân.
  - Bảng chuyển đổi:

16	2	16	2	16	2	16	2
0	0000	4	0100	8	1000	C	1100
1	0001	5	0101	9	1001	D	1101
2	0010	6	0110	A	1010	E	1110
3	0011	7	0111	B	1011	F	1111

Chuyển từ cơ số 2 sang cơ số b

- **Từ nhị phân sang bát phân:**
  - Nhóm từng bộ 3 bit trong biểu diễn nhị phân rồi chuyển sang kí số tương ứng trong hệ bát phân.
  - Bảng chuyển đổi:

8	2	8	2
0	000	4	100
1	001	5	101
2	010	6	110
3	011	7	111

Biểu diễn thông tin trên máy tính

Đơn vị lưu trữ thông tin

- Số học nhị phân sử dụng hai ký số 0 và 1.
- Bit (Binary Digit): Đơn vị chứa thông tin nhỏ nhất.
- Được lưu trong các thanh ghi hoặc trong các ô nhớ.
- Thanh ghi hoặc ô nhớ có kích thước 1 byte (8 bit) hoặc 1 word (16 bit)

Đơn vị lưu trữ thông tin

1 bit	0	2
2 bit	1 0	2^2
3 bit	2 1 0	2^3
n bit	n-1 … 5 4 3 2 1 0	2^n

$0\ldots000 \rightarrow 1\ldots111 = 2^n - 1$

Đơn vị lưu trữ thông tin

- Các đơn vị đo thông tin lớn hơn:

Tên gọi	Ký hiệu	Giá trị
Byte	B	8 bit
KiloByte	KB	2^{10} B = 1024 Byte
MegaByte	MB	2^{10} KB = 2^{20} Byte
GigaByte	GB	2^{10} MB = 2^{30} Byte
TeraByte	TB	2^{10} GB = 2^{40} Byte
Peta	PB	2^{10} TB = 2^{50} Byte

Biểu diễn số nguyên không dấu

- Đặc điểm
  - Biểu diễn các đại lương luôn dương.
  - Ví dụ: chiều cao, cân nặng, mã ASCII…
  - Tất cả bit được sử dụng để biểu diễn giá trị.
  - Số nguyên không dấu 1 byte lớn nhất là $1111\ 1111_2 = 2^8 - 1 = 255_{10}$.
  - Số nguyên không dấu 1 word lớn nhất là $1111\ 1111\ 1111\ 1111_2 = 2^{16} - 1 = 65535_{10}$.
  - Tùy nhu cầu có thể sử dụng số 2, 3… word.

Biểu diễn số nguyên có dấu

- Đặc điểm
  - Lưu các số dương hoặc âm.
  - Bit cao nhất dùng để biểu diễn dấu
  - Ví dụ đối với số nguyên sử dụng 8 bit lưu trữ (gọi tắt là số nguyên 8 bits)
    - 0 biểu diễn số dương. VD: 0101 0011
    - 1 biểu diễn số âm. VD: 1101 0011
  - Số âm trong máy được biểu diễn ở dạng số bù 2.

Số bù 1 và số bù 2

Số 5 (1 byte)		0	0	0	0	0	1	0	1
Số bù 1 của 5		1	1	1	1	1	0	1	0
+									1
Số bù 2 của 5		1	1	1	1	1	0	1	1
+ Số 5		0	0	0	0	0	1	0	1
Kết quả	1	0	0	0	0	0	0	0	0

Biểu diễn số nguyên có dấu

- Nhận xét
  - Số bù 2 của x cộng với x là một dãy toàn bit 0 (không tính bit 1 cao nhất do vượt quá phạm vi lưu trữ). Do đó số bù 2 của x chính là giá trị âm của x hay – x.
  - Đổi số thập phân âm –5 sang nhị phân?
    - => Đổi 5 sang nhị phân rồi lấy số bù 2 của nó.
  - Thực hiện phép toán a – b?
    - a – b = a + (–b) => Cộng với số bù 2 của b

Tính giá trị có dấu và không dấu

- Tính giá trị không dấu và có dấu của 1 số?
  - Ví dụ số word (16 bit): 1100 1100 1111 0000
  - Số nguyên không dấu ?
    - Tất cả 16 bit lưu giá trị.
    - => giá trị là 52464.
  - Số nguyên có dấu ?
    - Bit cao nhất = 1 do đó số này là số âm.
    - => độ lớn là giá trị của số bù 2.
    - Số bù 2 = 0011 0011 0001 0000 = 13072.
    - => giá trị là –13072.

Biểu diễn số thực

- Sử dụng dấu chấm động (floating-point).
- Ví dụ:
  - Số thực hệ 10: -123.4*d* = -12.34 x 10^1 = -1.234 x 10^2 = -0.1234 x 10^3
- Biểu diễn khoa học: -1.234 x 10^2
- Chia làm 3 phần:
  - 1 bit để biểu diễn dấu.
  - Một chuỗi bit để biểu diễn số mũ.
  - Một chuỗi bit để biểu diễn phần định trị

Biểu diễn số thực

- Với ví dụ trên: -1.234×10^2
  - Bit biểu diễn dấu là 1 (ứng với giá trị âm)
  - Biểu diễn số mũ là 2
  - Biểu diễn phần trị là 1234
    - Quy ước bên trái dấu chấm là 1 ký số khác không

Hệ thống mã hóa

Lưu trữ - Hiển thị

- Lưu trữ và xử lý: bit ~ số
- Hiển thị văn bản: ký tự # số

-> Cần phải có bảng mã, làm nhiệm vụ quy ước sự tương ứng giữa giá trị số và giá trị ký tự

- Bảng mã thông dụng: ASCII và Unicode

ASCII

- **A**merican **S**tandard **C**ode for **I**nformation **I**nterchange
- ASCII được công bố làm tiêu chuẩn lần đầu vào năm 1963
- Là bộ mã ký tự dựa trên bảng chữ cái tiếng Latinh (a – z, A – Z)
- Ban đầu, bảng mã ASCII chứa 128 mô tả cặp ký tự và số

ASCII – Các ký tự thông dụng

- Ký tự in được:
  - ‘ ’ (khoảng trắng): 32 (0x20)
  - ‘0’ -> ‘9’: 48 (0x30) -> 57 (0x39)
  - ‘A’ -> ‘Z’: 65 (0x41) -> 90 (0x5A)
  - ‘a’ -> ‘z’: 97 (0x61) -> 122 (0x7A)
- Ký tự điều khiển:
  - ký tự rỗng: 0
  - ‘ ’ (tab): 9
  - ký tự xuống dòng: 10
  - ký tự về đầu dòng: 13

ASCII

- Bảng mã ASCII mở rộng chứa 256 mô tả cặp ký tự số
  - 128 ký tự đầu giống ASCII ban đầu
  - 128 ký tự sau bao gồm 1 số ký hiệu tiếng Hy Lạp ('α', 'β', 'π', …), các biểu diễn tiền tệ ('£', '¥', …), …
- Bảng mã ASCII không thể biểu diễn các ký tự của các ngôn ngữ khác như tiếng Việt (có dấu), Nga, Nhật, …

Unicode

- **Unicode** là bộ mã chuẩn quốc tế được thiết kế để dùng làm bộ mã duy nhất cho tất cả các ngôn ngữ khác nhau trên thế giới
- Hiện tại, mã unicode có 1.114.112 mã (code points), được chia thành 17 miền, mỗi miền có 65535 (2^{16}) mã

Unicode

- Có nhiều cách biểu diễn mã unicode tùy theo kích thước lưu trữ mỗi phần tử mã (code point)
  - UTF – 8: sử dụng từ 1 -> 4 bytes
  - UTF – 16: sử dụng 2 bytes
  - UTF – 32: sử dụng 4 bytes

Unicode và Tiếng Việt

- Trong bảng mã unicode, có phần mã dành riêng tương ứng với những ký tự tiếng Việt
- Bảng mã unicode các ký tự tiếng Việt http://vietunicode.sourceforge.net/charset/v3.htm

Unicode - Font

- Mỗi bộ mã có nhiều cách hiển thị
- Font thực hiện chức năng ánh xạ giá trị mã unicode với hình [ký tự] hiển thị tương ứng
- Các font hỗ trợ unicode (có tiếng Việt) phổ biến:
  - Times New Roman,
  - Arial,
  - Tahoma,
  - …

Hệ thống tập tin

Phân loại tập tin

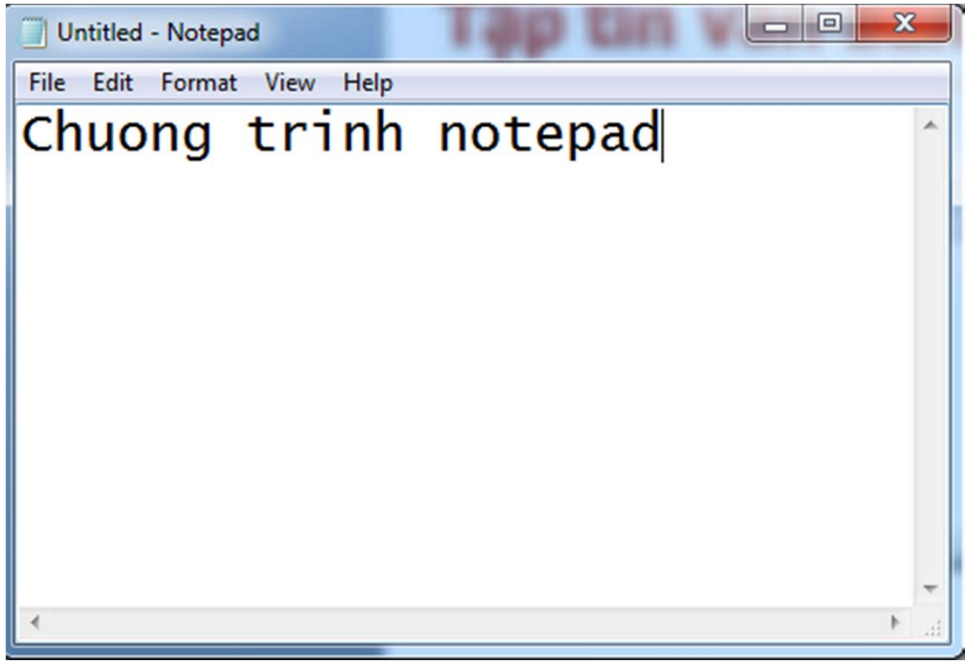
- Tập tin văn bản thô
- Tập tin nhị phân

Tập tin văn bản thô – Khái niệm

- Thường dùng để chỉ tập tin văn bản có cấu trúc đơn giản và thông dụng, có thể xem nội dung và sửa chữa bằng các lệnh của hệ điều hành hay chương trình soạn thảo đơn giản
- Không có header – phần thông tin nằm ở đầu tập tin, mô tả cấu trúc dữ liệu của tập tin sau phần header

Tập tin văn bản thô – Ví dụ

- *.txt
- *.cpp



Tập tin văn bản thô

- Tập tin theo cấu trúc ANSI text (hay ASCII text): chứa các ký tự (mã từ) trong bảng mã ASCII
- Đôi khi tập tin văn bản được tổ chức theo quy ước để phục vụ 1 mục đích xác định: gọi là tập tin văn bản được cấu trúc hóa

Tập tin văn bản thô ANSI text được cấu trúc

- Ví dụ: Cần lưu trữ nội dung ma trận có 3 dòng 4 cột
  - Dòng đầu cho biết số dòng, số cột
  - 3 dòng tiếp theo mỗi dòng 4 giá trị: nội dung ma trận
- Các loại tập tin văn bản cấu trúc thông dụng hiện tại:
  - *.RTF
  - *.HTML

Tập tin văn bản thô dạng mở rộng

- Văn bản thô ANSI text dựa trên cơ sở các ký tự 8-bit (256 ký hiệu)

-> Bất tiện khi lưu văn bản của nhiều ngôn ngữ

-> Văn bản thô dạng mở rộng cho phép lưu trữ được nhiều ngôn ngữ

  – Unicode text (lưu ký tự UTF-16)
  – UTF-8 text

Tập tin nhị phân

- Đa số tập tin nhị phân được cấu trúc hóa theo một quy ước nào đó
- Thường có phần header: chứa thông tin mô tả sự bố trí và mối liên hệ của các byte dữ liệu ở phía sau
- Mở bằng các công cụ (phần mềm) chuyên dụng

Tập tin nhị phân thông dụng

- Tập tin mã thực thi:
  - *.EXE, *.COM, *.DLL trên Windows
- Tập tin văn bản tích hợp dữ liệu văn bản, hình ảnh, bảng biểu
  - *.DOC của MS Word hay Open Office
- Tập tin multimedia
  - Ảnh: *.bmp, *.jpg, …
  - Âm thanh: *.wav, *.mp3, …
  - Video: *.avi, *.mp4, …