

# MÔN: PHÁT TRIỂN HỆ THỐNG ĐA PHƯƠNG TIỆN

Hồ Phúc Lâm - LT02 – 12/08/2024

Chủ đề: Unicode

GVHD: Nguyễn Thành Thái

## 1) Ôn tập hệ thập lục phân, nhị phân

Thập phân	Nhị phân (4bits)	Thập lục (hex)
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

## 2) Chuyển thập phân sang nhị phân dec, hex

$2^0$	$2^1$	$2^2$	$2^3$	$2^4$	$2^5$	$2^6$	$2^7$	$2^8$	$2^9$
1	2	4	8	16	32	64	128	256	512

a)  $78_{(10)} = 64 + 8 + 4 + 2 = 2^6 + 2^3 + 2^2 + 2^1 \Rightarrow 0100\ 1110_{(2)} \Rightarrow 4E_{(16)}$

b)  $94_{(10)} = 64 + 16 + 8 + 4 + 2 = 2^6 + 2^4 + 2^3 + 2^2 + 2^1 \Rightarrow 0101\ 1110 \Rightarrow 5E_{(16)}$

c)  $134_{(10)} = 128 + 4 + 2 = 2^7 + 2^2 + 2^1 \Rightarrow 1000\ 0110 \Rightarrow 86_{(16)}$

## 3) Cho code point U+0061, hãy tính mã UTF-8 1 byte

--chuyển mã hex to binary nhị phân

Bước 1: chuyển mã code point  $\Rightarrow$  chuỗi nhị phân

Code point U+0061 => 61 => 0110 0001

Bước 2: Chuyển 7 bits thấp (bên phải) của code point vào định dạng UTF-8 (bên trái)

0110 0001

0	1	1	0	0	0	0	1
---	---	---	---	---	---	---	---

Kết quả: 61

**4a) Cho code point U+0110, hãy tính mã UTF-8 2bytes và cho biết ký tự biểu diễn (mã hóa) trên các công cụ.**

Bước 1: viết code point dưới dạng chuỗi nhị phân

U+0110 => 0110 => 0000 0001 0001 0000<sub>(2)</sub>

--lấy ra 11 bit từ bên phải sang trái

0000 0001 0001 0000 => 001 0001 0000

--chia ra làm đôi để tách ra làm 2 phần

Phần 1 đầu tiên (5 bit cao) xét bên trái

Phần 2 sau (6 bit thấp) xét bên phải

Bước 2: đưa xuống bảng

1	1	0	0	0	1	0	0
1	0	0	1	0	0	0	0

--110 dòng I và 10 dòng II là cố định

Bước 3: Đáp số:

1100 0100

1001 0000

C 4

9 0

Đáp số: C4 90 => Đ (ký tự biểu diễn)

**4b) Cho code point U+0102, hãy tính mã UTF-8 2bytes và cho biết ký tự biểu diễn (mã hóa) trên các công cụ.**

Bước 1: viết code point dưới dạng chuỗi nhị phân

U+0102 => 0102 => 0000 0001 0000 0010<sub>(2)</sub>

--lấy ra 11 bit từ bên phải sang trái

0000 0001 0000 0010 => 001 0000 0010

--chia ra làm đôi để tách ra làm 2 phần

Phần 1 đầu tiên (5 bit cao) xét bên trái

Phần 2 sau (6 bit thấp) xét bên phải

Bước 2: đưa xuống bảng

1	1	0	0	0	1	0	0
1	0	0	0	0	0	1	0

Bước 3: Đáp số: C4 82 => Æ (ký tự biểu diễn)

**4c) Cho code point U+01A9, hãy tính mã UTF-8 2bytes và cho biết ký tự biểu diễn (mã hóa) trên các công cụ.**

Bước 1: viết code point dưới dạng chuỗi nhị phân

U+01A9 => 01A9 => 0000 0001 1010 1001

--lấy ra 11 bit từ bên phải sang trái

0000 0001 1010 1001 => 001 1010 1001

--chia ra làm đôi để tách ra làm 2 phần

Phần 1 đầu tiên (5 bit cao) xét bên trái

Phần 2 sau (6 bit thấp) xét bên phải

Bước 2: đưa xuống bảng

1	1	0	0	0	1	1	0
1	0	1	0	1	0	0	1

Bước 3: Đáp số:

1100 0110

1010 1001

C 6

A 9

**Đáp số: C6 A9 => Σ (kí tự biểu diễn)**

**5) Cho code point U+0E10, hãy tính mã UTF-8 3 bytes và cho biết ký tự biểu diễn (mã hóa) trên các công cụ.**

Bước 1: U+0E10 chuyển => 0E10 => 0000 1110 0001 0000<sub>(2)</sub>

--Lấy 16 bit từ phải sang:

0000 1110 0001 0000

--Chia 3 phần Gồm I:4bits và II:6bits và III:6bits

0000 1110 0001 0000

Bước 2: chèn vào bảng

1	1	1	0	0	0	0	0
1	0	1	1	1	0	0	0
1	0	0	1	0	0	0	0

**Bước 3: chuyển đổi => E0 B8 90 = ๐ (THAI CHARACTER THO THAN)**

**6) Cho code point U+1FA77, hãy tính mã UTF-8 4 bytes và cho biết ký tự biểu diễn (mã hóa) trên các công cụ.**

Bước 1: chuyển đổi nhị phân

U+1FA77 => 1FA77 = 0000 0001 1111 1010 0111 0111<sub>(2)</sub>

--Lấy ra 21 bits của code point

0000 0001 1111 1010 0111 0111

--Chia 4 phần I: 3 bít, II 6bít, III 6bít, IV 6bít

0 0001 1111 1010 0111 0111

Bước 2: Chuyển vào bảng

1	1	1	1	0	0	0	0
1	0	0	1	1	1	1	1
1	0	1	0	1	0	0	1
1	0	1	1	0	1	1	1

Bước 3:

**F0 9F A9 B7 => PINK HEART**