

**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI**

**KHOA CÔNG NGHỆ THÔNG TIN**

----o0o----



**BÁO CÁO BÀI TẬP LỚN**

**NHẬP MÔN CÔNG NGHỆ THÔNG TIN**

***Đề tài: Tóm tắt các thao tác cơ bản trên hệ đếm thông dụng.***

***Giới thiệu các hệ thống mã hóa phổ biến.***

***Lớp: Khoa học máy tính – K63***

***Sinh viên thực hiện:***

***Nguyễn Tú Anh – 223630670***

***Dương Trung Hiếu – 223630690***

***Thân Văn Quân – 223630706***

***Nguyễn Xuân Quang – 223630705***

***Hà Văn Trọng – 223630717***

**Hà Nội, tháng 10 năm 2022**

**MỤC LỤC**

Mục lục…………………………………………………………………………2

[DANH SÁCH HÌNH ẢNH 4](#_Toc118594836)

[LỜI NÓI ĐẦU 5](#_Toc118594837)

[PHẦN MỞ ĐẦU 6](#_Toc118594838)

[PHẦN I: CÁC THAO TÁC CƠ BẢN TRÊN HỆ ĐẾM THÔNG DỤNG 9](#_Toc118594839)

[1.1 Hệ đếm 9](#_Toc118594840)

[1.2 Các hệ đếm thông dụng 9](#_Toc118594841)

[***1.2.1*** ***Hệ nhị phân (Binary System)*** 9](#_Toc118594842)

[***1.2.2*** ***Hệ thập phân (Decimal System)*** 9](#_Toc118594843)

[***1.2.3*** ***Hệ thập lục phân (Hexadecimal System)*** 9](#_Toc118594844)

[1.3 Các thao tác cơ bản trên hệ đếm thông dụng 10](#_Toc118594845)

[***1.3.1*** ***Chuyển đổi hệ cơ số b sang hệ cơ số 10*** 10](#_Toc118594846)

[1.3.1.1 Chuyển đổi hệ cơ số 2 sang hệ cơ số 10 10](#_Toc118594847)

[1.3.1.2 Chuyển đổi hệ cơ số 16 sang hệ cơ số 10 10](#_Toc118594848)

[***1.3.2*** ***Chuyển đổi hệ cơ số 10 sang hệ cơ số b*** 10](#_Toc118594849)

[1.3.2.1 Chuyển đổi hệ cơ số 10 sang hệ cơ số 2 10](#_Toc118594850)

[1.3.2.2 Chuyển đổi hệ cơ số 10 sang hệ cơ số 16 12](#_Toc118594851)

[***1.3.3*** ***Chuyển đổi giữa các hệ cơ số 2, hệ cơ số 16*** 12](#_Toc118594852)

[1.3.3.1 Từ hệ cơ số 2 sang hệ cơ số 16 12](#_Toc118594853)

[1.3.3.2 Từ hệ cơ số 16 sang hệ cơ số 2 12](#_Toc118594854)

[**1.3.4** **Các phép tính toán cơ bản trên hệ thập phân** 13](#_Toc118594855)

[1.3.4.1 Phép cộng 13](#_Toc118594856)

[1.3.4.2 Phép trừ 13](#_Toc118594857)

[1.3.4.3 Phép nhân 13](#_Toc118594858)

[1.3.4.4 Phép chia 14](#_Toc118594859)

[**1.3.5** **Các phép tính toán cơ bản trên hệ nhị phân** 14](#_Toc118594860)

[1.3.5.1 Phép cộng 14](#_Toc118594861)

[1.3.5.2 Phép trừ 15](#_Toc118594862)

[1.3.5.3 Phép nhân 15](#_Toc118594863)

[1.3.5.4 Phép chia 15](#_Toc118594864)

[**1.3.6** **Các phép tính toán cơ bản trên hệ thập lục phân** 16](#_Toc118594865)

[1.3.6.1 Phép cộng 16](#_Toc118594866)

[1.3.6.2 Phép trừ 16](#_Toc118594867)

[1.3.6.3 Phép nhân 16](#_Toc118594868)

[1.3.6.4 Phép chia 16](#_Toc118594869)

[PHẦN II : CÁC HỆ THỐNG MÃ HOÁ PHỔ BIẾN 17](#_Toc118594870)

[2.1 Hệ thống mã hoá 17](#_Toc118594871)

[**2.1.1** **Mã hoá và giải mã** 17](#_Toc118594872)

[2.1.1.1 Khái niệm về mã hoá và giải mã: 17](#_Toc118594873)

[2.1.1.2 Các thuật toán mã hoá cơ bản: 18](#_Toc118594874)

[**2.1.2** **Hệ thống mã hoá** 20](#_Toc118594875)

[2.2 Các hệ thống mã hoá cơ bản 21](#_Toc118594876)

[**2.2.1** **Hệ thống mã hoá ASCII** 21](#_Toc118594877)

[2.2.1.1 Tổng quan về ASCII: 21](#_Toc118594878)

[2.2.1.2 Lịch sử ra đời và phát triển của ASCII [4] 21](#_Toc118594879)

[2.2.1.3 Các kí tự trong bảng mã ASCII tiêu chuẩn: 22](#_Toc118594880)

[2.2.1.4 Bảng mã ASCII mở rộng 23](#_Toc118594881)

[**2.2.2** **Hệ thống mã hóa Unicode** 24](#_Toc118594882)

[2.2.2.1 Sơ lược về bảng mã Unicode 24](#_Toc118594883)

[2.2.2.2 Kho chữ của hệ thống mã hóa Unicode 25](#_Toc118594884)

[2.2.2.3 Các bảng mã của Unicode 26](#_Toc118594885)

[2.2.2.4 Cách bảng mã Unicode mã hoá hiển thị tiếng Việt 26](#_Toc118594886)

[KẾT LUẬN 28](#_Toc118594887)

[TÀI LIỆU THAM KHẢO 29](#_Toc118594888)

# **DANH SÁCH HÌNH ẢNH**

[Hình 1.1 Đổi phần nguyên của 1 số từ hệ thập phân sang hệ nhị phân 9](#_Toc117920408)

[Hình 1.2 Đổi phần thập phân của một số từ hệ thập phân sang hệ nhị phân 10](#_Toc117920409)

[Hình 1.3 Đổi phần nguyên của một số từ hệ cơ số 10 sang hệ cơ số 16 11](#_Toc117920410)

HÌNH 1.4 PHÉP CHIA TRONG HỆ THẬP PHÂN .............................................. 13

HÌNH 1.5 PHÉP CHIA TRONG HỆ NHỊ PHÂN .................................................. 14

[Hình 2.1 Mã hoá 16](#_Toc117920411)

[Hình 2.2 Kiểm tra mật khẩu bằng hàm băm 18](#_Toc117920412)

[Hình 2.3 Quy trình mã hoá cơ bản 20](#_Toc117920413)

[Hình 2.4 Bảng mã ASCII - Các kí tự điều khiển 21](#_Toc117920414)

[Hình 2.5 Bảng mã ASCII mở rộng 23](#_Toc117920415)

# **LỜI NÓI ĐẦU**

Máy tính đã trải qua một chặng đường dài trong thời gian tương đối ngắn. Cùng với sự phát triển không ngừng về kỹ thuật máy tính và mạng điện tử, công nghệ thông tin cũng đã đạt được nhiều thành tựu to lớn, góp công trong công cuộc xây dựng và phát triển nhân loại.

Như chúng ta đã biết, máy tính được vận hành và lập trình bằng một số loại ngôn ngữ mã hóa. Phần cứng của máy tính cũng như phần mềm hoạt động trên một danh sách tập hợp các kí tự được xác định và mỗi kí tự được xác định bằng một số.

Do đó, nhóm chúng em đã lựa chọn đề tài: “Các thao tác cơ bản trên hệ đếm thông dụng và giới thiệu các hệ thống mã hóa phổ biến”.

# **PHẦN MỞ ĐẦU**

**Lí do chọn đề tài:**

Trong cuộc sống hàng ngày, con người giao tiếp với nhau thông qua một hệ thống ngôn ngữ quy ước, nhưng trong máy tính và các hệ thống số chỉ xử lí các dữ liệu nhị phân. Để máy tính thực hiện được những bài toán do con người tạo ra, người ta đặt ra vấn đề về mã hóa dữ liệu, giúp biến đổi những kí hiệu quen thuộc của con người sang những kí hiệu quen thuộc của máy tính.

**Mục tiêu đề tài:**

Đề tài với mục tiêu giúp mọi người hiểu về các thao tác cơ bản trên hệ đếm thông dụng (nhị phân, thập phân, thập lục phân) và một số hệ thống mã hóa phổ biến (ASCII và UNICODE).

**Nhiệm vụ nghiên cứu:**

**Phần 1:** Tóm tắt các thao tác cơ bản trên hệ đếm thông dụng.

**Phần 2:** Trình bày định nghĩa, ứng dụng của hệ thống mã hóa. Giới thiệu các hệ thống mã hóa phổ biến (ASCII, UNICODE).

**Đối tượng nghiên cứu:**

Hệ đếm thông dụng và một số hệ thống mã hóa phổ biến.

**BÁO CÁO PHÂN CÔNG NHIỆM VỤ VÀ KẾT QUẢ CỦA TỪNG THÀNH VIÊN TRONG NHÓM 5**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STT | Họ và tên | Mã SV | Nhiệm vụ | Kết quả | Chữ ký |
| 1 | Nguyễn Tú Anh | 223630670 | Sửa báo cáo, làm slides thuyết trình | Hoàn thành |  |
| 2 | Dương Trung Hiếu | 223630690 | Chuẩn bị phần 1 bài báo cáo | Hoàn thành |  |
| 3 | Thân Văn Quân | 223630706 | Chuẩn bị phần 2, hoàn thiện bài báo cáo | Hoàn thành |  |
| 4 | Nguyễn Xuân Quang | 223630705 | Tìm kiếm hình ảnh, tài liệu | Hoàn thành |  |
| 5 | Hà Văn Trọng | 223630717 | Làm slides thuyết trình | Hoàn thành |  |

**BÁO CÁO TIÉN ĐỘ HOÀN THÀNH BÀI TẬP LỚN**

1. **Kế hoạch thực hiện đề tài**

|  |  |  |
| --- | --- | --- |
| Nội dung | Thời hạn dự kiến | Kết quả |
| Các bước chuẩn bị ban dầu | 14 ngày | Hoàn thành |
| Tổng hợp thông tin | 31/10/2022 | Hoàn thành |
| Viết báo cáo | 31/10/2022 | Hoàn thành |
| Làm slide thuyết trình | 02/11/2022 | Hoàn thành |
| Tổng duyệt nội dung | 05/10/2022 | Hoàn thành |

1. **Theo dõi tiến độ hoạt động của các thành viên**
   1. Các bước chuẩn bị ban đầu

|  |  |  |
| --- | --- | --- |
| Nội dung | Ngày thực hiện | Kết quả |
| Bầu nhóm trưởng, thư kí | 26/10/2022 | Nhóm trưởng: Nguyễn Tú Anh  Thư kí: Hà Văn Trọng |
| Thống nhất dàn ý chung cho đề tài | Hoàn thành |
| Phân công công việc đến từng thành viên, thống nhất hạn nộp | Hoàn thành |

* 1. Tổng hợp thông tin

|  |  |  |  |
| --- | --- | --- | --- |
| Nội dung | Người thực hiện | Hạn nộp | Kết quả |
| Tổng hợp cơ sở lý thuyết, tài liêu tham khảo cho phần 1 | Dương Trung Hiếu | 28/10/2022 | Hoàn thành |
| Tổng hợp cơ sở lý thuyết, tài liệu tham khảo cho phần 2 | Thân Văn Quân | 28/10/2022 | Hoàn thành |
| Tổng hợp hình ảnh cho slide | Hà Văn Trọng | 30/10/2022 | Hoàn thành |

* 1. Viết báo cáo
  2. Làm silde thuyết trình
  3. Tổng duyệt nội dung

|  |  |
| --- | --- |
| Thư kí  (kí và ghi họ tên) | Nhóm trưởng  (kí và ghi họ tên) |
| Hà Văn Trọng | Nguyễn Tú Anh |

# **PHẦN I: CÁC THAO TÁC CƠ BẢN TRÊN HỆ ĐẾM THÔNG DỤNG**

## Hệ đếm

* Khái niệm: Là tập hợp các kí hiệu và qui tắc để biểu diễn và xác định giá trị các số.
* Mỗi hệ đếm có một số ký tự/số (ký số) hữu hạn. Tổng số ký số của mỗi hệ đếm được gọi là cơ số (base hay radix), ký hiệu là b.
* Ví dụ: Trong hệ đếm cơ số 10, dùng 10 kí tự là các chữ số từ 0 đến 9.
* Về mặt toán học, ta có thể biểu diễn một số theo hệ đếm cơ số bất kì.

## Các hệ đếm thông dụng

### ***Hệ nhị phân (Binary System)***

* Hệ nhị phân (hay hệ đếm cơ số hai) là một hệ đếm dùng hai kí tự để biểu đạt một giá trị số, bằng tổng số các lũy thừa của hai. Hệ đếm chỉ dùng 2 chữ số (0 và 1).
* Chữ số nhị phân gọi là **bit** (**bi**nary dig**t**). Ví dụ bit 0, bit 1.
* Bit là đơn vị thông tin nhỏ nhất.
* Giả sử số A được biểu diễn theo hệ nhị phân:

A(2) = anan-1…a1a0a-1a-2…a-m ai ∈ {0;1}

⇒ A(10) = an2n + an-12n-1 + … + a121 + a020 + a-12-1 + … + a-m2-m

* Ví dụ: 1000(2) = 1.23 + 0.22 +0.21 +0.20 = 8(10)

### ***Hệ thập phân (Decimal System)***

* Hệ đếm thập phân hay hệ đếm cơ số 10 bao gồm 10 kí số theo kí hiệu sau:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9

* Giả sử số A được biểu diễn dưới dạng:

A(10) = anan −1 . . . a1a0a−1 a−2 . . . a−m

⇒ A(10) = an10n + an-110n-1 + … + a1101 + a0100+ a-110-1 + … + a-m10-m

* Ví dụ: 279(10) = 2.102 + 7.101 + 9.100

### ***Hệ thập lục phân (Hexadecimal System)***

* Hệ thập lục phân là hệ đếm có 16 kí tự, từ 0 đến 9, từ A đến F (chữ hoa và chữ thường như nhau). Trong đó, A***⟺***10, B***⟺***11, C***⟺***12, D***⟺***13, E***⟺***14, F***⟺***15.
* Số thập lục phân có dạng: A(16)=anan-1an-2…a0**.**a­-1a-2…a-m

⇒ A(10) = an16n+an-116n-1+an-216n-2+…+a0160+a-116-1+a-216-2+…+a-m16-m

* Ví dụ: 3CE(16) = 3.162 + 12.161 + 14.160= 974(10)

## Các thao tác cơ bản trên hệ đếm thông dụng

### ***Chuyển đổi hệ cơ số b sang hệ cơ số 10***

A(10) = anbn + an-1bn-1 +…+ a0b0 + a-1b-1 +…+ a-mb-m

Ví dụ: 1000(2) = 1.23 + 0.22 + 0.21 + 0.20 = 8(10)

#### Chuyển đổi hệ cơ số 2 sang hệ cơ số 10

A(10) = an2n + an-12n-1 +…+ a020 + a-12-1 +…+ a-m2-m

Ví dụ: 1101(2) = 1.23 +1.22 + 0.21 + 1.20 = 13(10)

#### Chuyển đổi hệ cơ số 16 sang hệ cơ số 10

A(10) = an16n + an-116n-1 +…+ a0160 + a-116-1 +…+ a-m16-m

Ví dụ: 3CE(16) = 3.162 + 12.161 + 14.160= 974(10)

### ***Chuyển đổi hệ cơ số 10 sang hệ cơ số b***

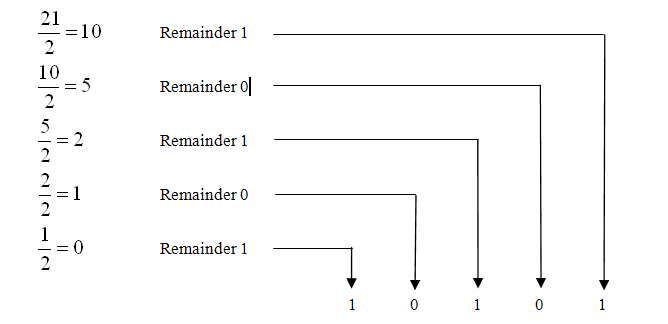
*Bước 1:* Đổi phần nguyên: Lấy số nguyên thập phân N(10) lần lượt chia cho b cho đến khi thương số bằng 0. Kết quả số chuyển đổi M(b) là các số dư trong phép chia được viết ra theo thứ tự ngược lại.

*Bước 2:*  Đổi phần lẻ: Lấy phần thập phân (sau dấu phẩy) N(10) lần lượt nhân với b cho đến khi phần thập phân của tích số bằng 0. Kết quả số chuyển đổi M(b) là các số phần nguyên trong phép nhân được viết ra theo thứ tự phép tính.

#### Chuyển đổi hệ cơ số 10 sang hệ cơ số 2

Bước 1. Đổi phần nguyên của số đó từ hệ thập phân sang hệ nhị phân: Chia liên tiếp phần nguyên cho 2 cho đến khi thương bằng 0

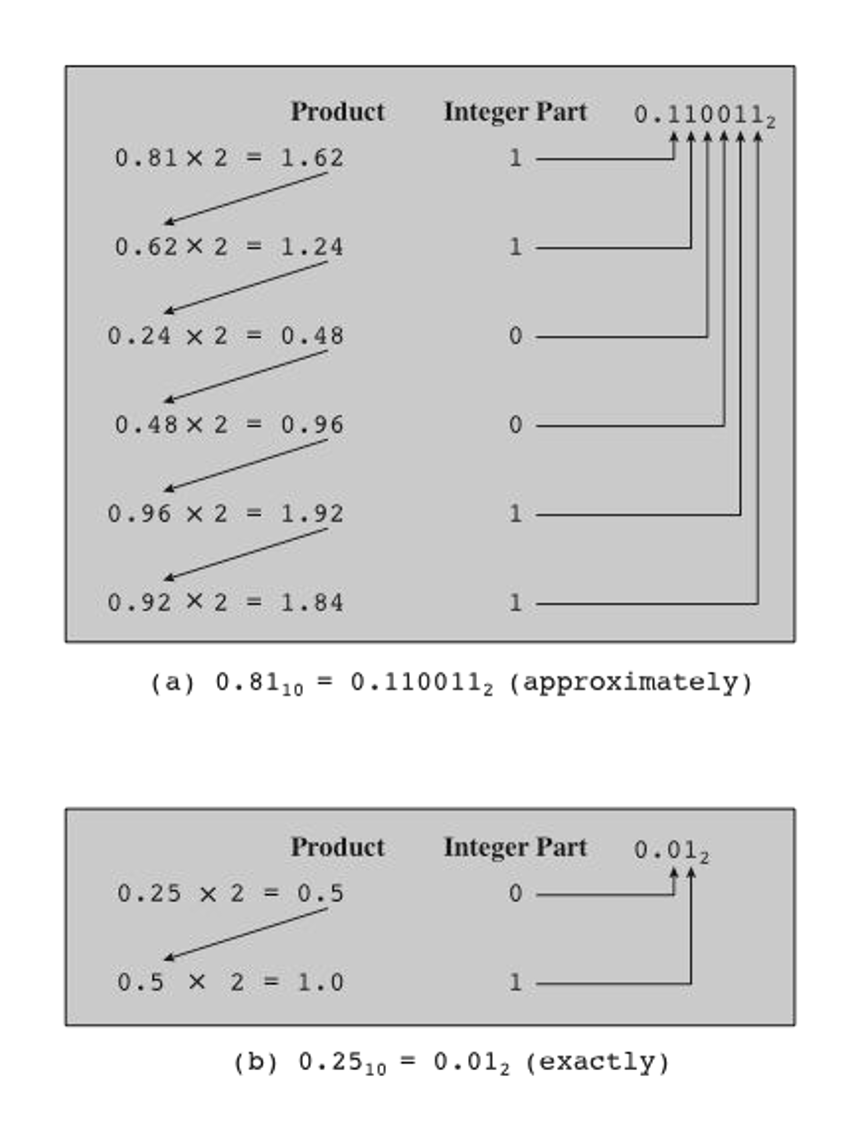
→kết quả chuyển đổi là dãy số dư liên tiếp viết theo thứ tự ngược lại.



Hình 1. 1 Đổi phần nguyên của 1 số từ hệ thập phân sang hệ nhị phân

Bước 2. Đổi phần thập phân của số đó từ hệ thập phân sang hệ nhị phân: Nhân liên tiếp phần lẻ với 2 cho đến khi phần lẻ bằng 0

→ kết quả chuyển đổi là dãy phần nguyên trong mỗi phép nhân được viết ra theo thứ tự tính toán.

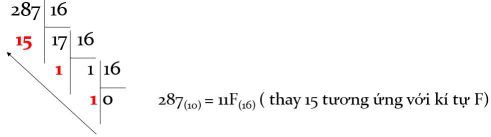


Hình 1. 2 Đổi phần thập phân của một số từ hệ thập phân sang hệ nhị phân

#### Chuyển đổi hệ cơ số 10 sang hệ cơ số 16

Bước 1. Đổi phần nguyên của số đó từ hệ thập phân sang hệ thập lục phân: Chia liên tiếp phần nguyên cho 16 cho đến khi thương bằng 0

→kết quả chuyển đổi là dãy số dư liên tiếp viết theo thứ tự ngược lại.



Hình 1.3 Đổi phần nguyên của một số từ hệ cơ số 10 sang hệ cơ số 16

Bước 2. Đổi phần thập phân của số đó từ hệ thập phân sang hệ nhị phân: Nhân liên tiếp phần lẻ với 16 cho đến khi phần lẻ bằng 0

→ kết quả chuyển đổi là dãy phần nguyên trong mỗi phép nhân được viết ra theo thứ tự tính toán.

### ***Chuyển đổi giữa các hệ cơ số 2, hệ cơ số 16***

#### Từ hệ cơ số 2 sang hệ cơ số 16

Quy tắc: Tách số ở hệ cơ số 2 thành các dãy 4 bit liền nhau từ phải qua trái sau đó nhân các số tách được từ phải qua trái lần lượt với 20, 21, 22, 23; kết quả của mỗi phép nhân sẽ tương ứng với 1 số trong hệ cơ số 16.

Lưu ý: Nếu có dãy tách ra bị thiếu bit thì thêm trước dãy đó số 0

Ví dụ: 00111101(2) = ?(16)

1101 = 1.20 + 0.21 + 1.22 + 1.23 = 13

0011 = 1.20 + 1.21 + 0.22 + 0.23 = 3

00111101(2) = 3D(16)

#### Từ hệ cơ số 16 sang hệ cơ số 2

Quy tắc: Lấy từng số N(16) lần lượt chia với 2 cho đến khi thương số bằng 0. Kết quả số chuyển đối M(2) là các số dư trong phép chia được viết ra theo thứ tự ngược lại. Một chữ số trong hệ cơ số 16 sẽ tương ứng với dãy 4 bit liền nhau trong hệ cơ số 2.

Lưu ý: Nếu ghép các số dư trong phép chia trên không đủ 4 số thì thêm trước đó số 0

Ví dụ: D30(16) = ?(2)

 dư 1,  dư 0,  dư 1,  dư 1  → D(16)= 1101(2)

 dư 1,  dư 1 → 3(16)= 0011(2)

 dư 0 → 0(16)= 0000(2)

→ D30(16) = 110100110000(2)

### **Các phép tính toán cơ bản trên hệ thập phân**

#### Phép cộng

Quy tắc: - Viết số hạng này dưới số hạng kia làm sao cho các chữ số ở cùng một hàng đặt thẳng cột với nhau.

- Cộng như cộng các số tự nhiên.

- Viết dấu phẩy ở tổng thẳng cột với các dấu phẩy của các số hạng.

Ví dụ: 

#### Phép trừ

Quy tắc: - Viết số trừ dưới số bị trừ sao cho các chữ số ở cùng một hàng đặt thẳng cột nhau.

- Thực hiện phép trừ như trừ các số tự nhiên.

- Viết dấu phẩy ở hiệu thẳng cột với các dấu phẩy của số bị trừ và số trừ.

Ví dụ: 

#### Phép nhân

Quy tắc: - Nhân như nhân các số tự nhiên.

- Đếm xem trong phần thập phân của cả hai thừa số có bao nhiêu chữ số rồi dùng dấu phẩy tách ở tích ra bấy nhiêu chữ số kể từ phải sang trái.

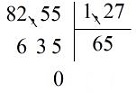
Ví dụ: 

#### Phép chia

Quy tắc: - Đếm xem có bao nhiêu chữ số ở phần thập phân của số chia thì chuyển dấu phẩy ở số bị chia sang bên phải bấy nhiêu chữ số.

- Bỏ dấu phẩy ở số chia rồi thực hiện phép chia như chia số tự nhiên.

Ví dụ: 



*Hình 1.4 Phép chia trong hệ thập phân*

### **Các phép tính toán cơ bản trên hệ nhị phân**

#### Phép cộng

Quy tắc: Giống như quy tắc cộng hai số thập phân

Lưu ý: 0 + 0 = 0

0 + 1 = 1

1 + 0 = 1

1 + 1 = 0 nhớ 1 (đem qua bít cao hơn).

Ngoài ra nếu cộng nhiều số nhị phân cùng một lúc ta nên nhớ :

- Nếu số bit 1 chẵn, kết quả là 0;

- Nếu số bit 1 lẻ kết quả là 1

- Và cứ 1 cặp số 1 cho 1 số nhớ (bỏ qua số 1 dư, thí dụ với 5 số 1 ta kể là 2 cặp)

Ví dụ: 

#### Phép trừ

Quy tắc: Giống như quy tắc trừ hai số thập phân

Lưu ý: 0 – 0 = 0

1 – 1 = 0

1 – 0 = 1

0 – 1 = 1 (nhớ 1 cho bit cao hơn)

Ví dụ: 

#### Phép nhân

Quy tắc: Giống như quy tắc nhân hai số thập phân

Lưu ý: 0 x 0 = 0

0 x 1 = 0

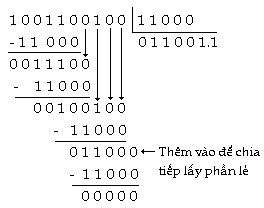
1 x 1 = 1

Ví dụ: 

#### Phép chia

Quy tắc: Giống như quy tắc chia hai số thập phân

Ví dụ:



*Hình 1.5 Phép chia trong hệ nhị phân*

### **Các phép tính toán cơ bản trên hệ thập lục phân**

#### Phép cộng

Quy tắc: Giống như quy tắc cộng hai số thập phân

Lưu ý: Số mượn là 16

Ví dụ:



#### Phép trừ

Quy tắc: Giống như quy tắc trừ hai số thập phân

Lưu ý: Số mượn là 16

Ví dụ: 

#### Phép nhân

Quy tắc: Giống như quy tắc nhân hai số thập phân

Lưu ý: Sau khi nhân hai số ( VD: E x 3= 42 thì phải chuyển số 42(10) = 2A(16) ( 2 là số nhớ vào B ) ).

Ví dụ: 

#### Phép chia

Quy tắc: Chuyển các số từ hệ thập lục phân sang hệ thập phân rồi thực hiện phép chia, sau khi thực hiện phép chia thì lấy kết quả thu được chuyển sang hệ thập lục phân

|  |  |
| --- | --- |
| Ví dụ: | → 3E8(16) = 1000(10)  → 14(16) = 20(10)  →  → 50(10)= 32(16)  → |

# **PHẦN II : CÁC HỆ THỐNG MÃ HOÁ PHỔ BIẾN**

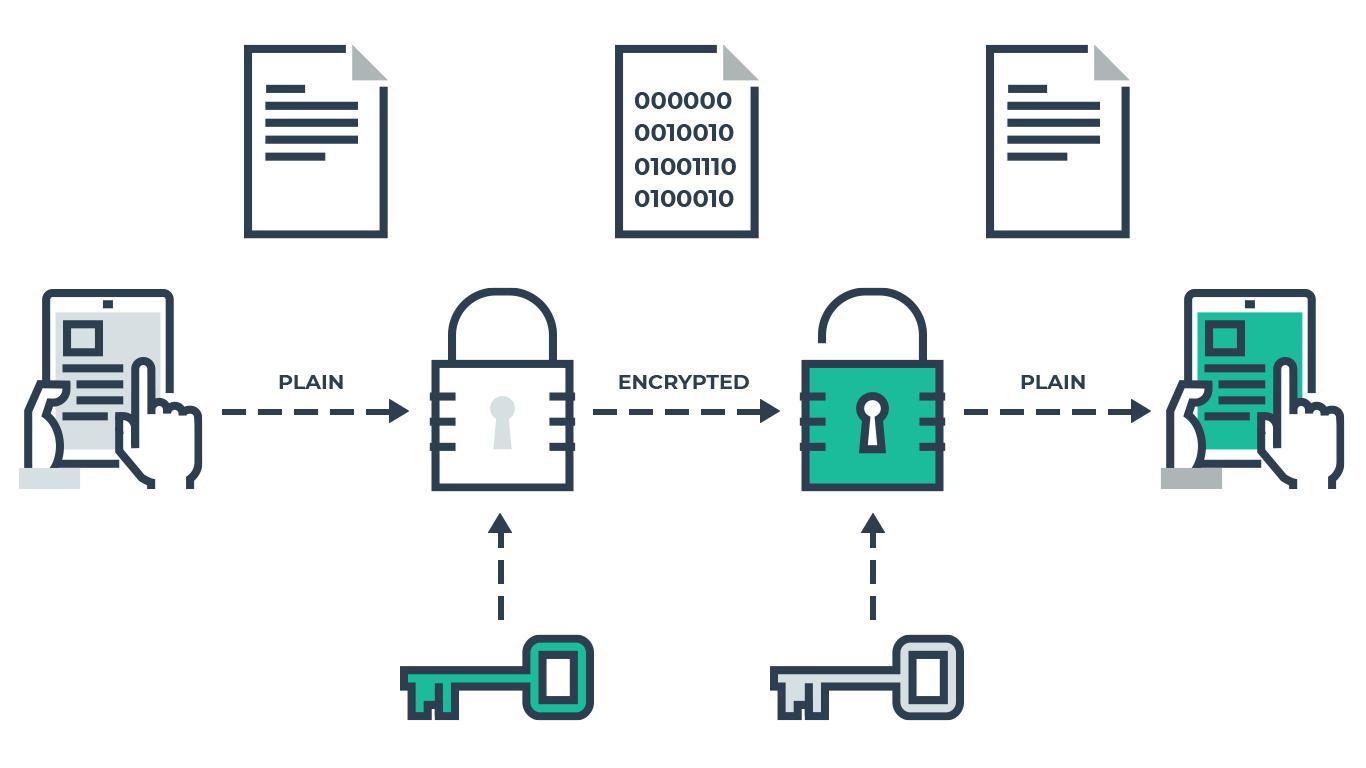
## Hệ thống mã hoá

### **Mã hoá và giải mã**

Từ xưa con người đã có ý thức về việc truyền thông tin một cách bí mật. Con người thời nguyên thủy khi săn bắt đã biết dùng ám hiệu để thông báo cho nhau. Và càng ngày thì việc bảo mật thông tin càng trở nên cấp thiết. Một ví dụ là máy Enigma được phát minh đầu tiên bởi kỹ sư người Đức Arthur Scherbius vào giai đoạn cuối của Chiến tranh thế giới thứ nhất được sử dụng cho mục đích dân sự và được bán thương mại.[[1]](https://vi.m.wikipedia.org/wiki/M%C3%A1y_Enigma) Nhưng sau đó máy Enigma chủ yếu được áp dụng cho những mục đích quân sự. Quân đội Đức Quốc xã đã sử dụng và cải tiến máy Enigma để mã hoá các thông tin và các bức điện nhằm phục vụ cho Chiến tranh thế giới thứ hai. Và đặc biệt cho đến hiện nay chúng ta đang sống trong môi trường Internet vô cùng mở, những thông tin mà ta gửi lên hoặc nhận về đều có nguy cơ bị đánh cắp và lợi dụng cho mục đích xấu. Vậy nên mã hoá dữ liệu được tạo ra nhằm bảo vệ con người trong môi trường mạng

#### Khái niệm về mã hoá và giải mã:

* Trong ngành mật mã học – một ngành toán học ứng dụng cho công nghệ thông tin, mã hóa là phương pháp để biến thông tin (phim ảnh, văn bản, hình ảnh...) từ định dạng bình thường sang dạng thông tin không thể hiểu được nếu không có phương tiện giải mã.

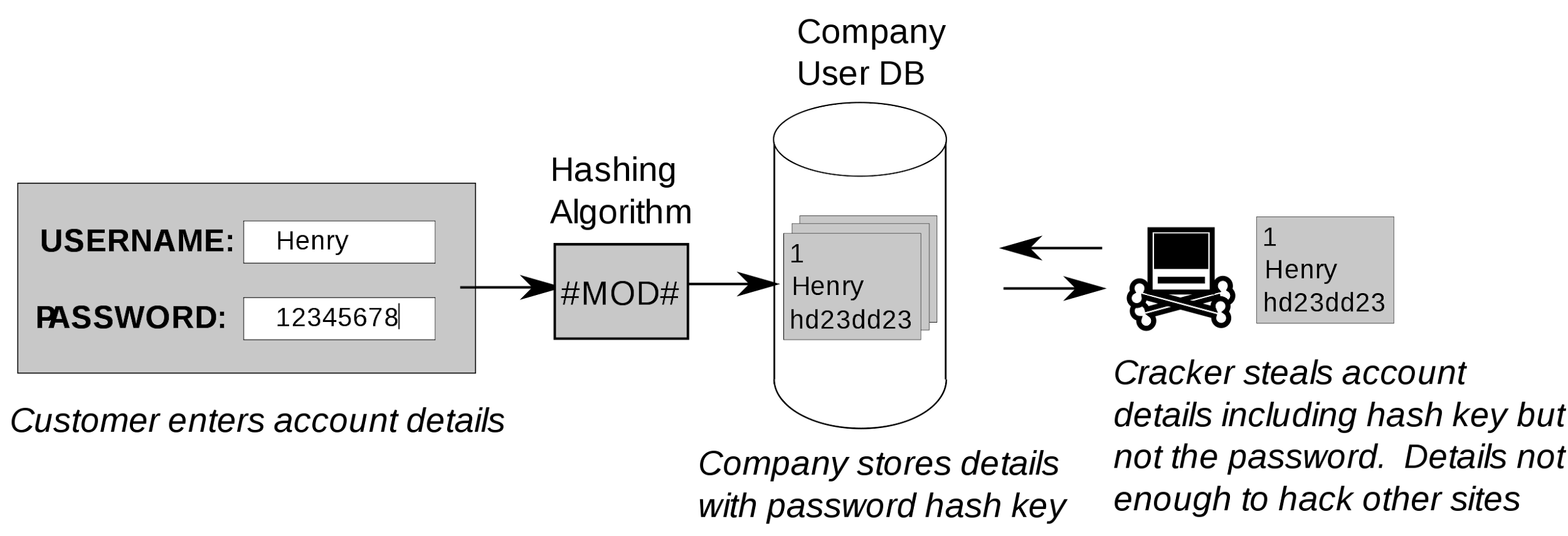
[](https://www.techsignin.com/khai-niem-ve-ma-hoa/)

Hình 2.1 Mã hoá

* Giải mã là phương pháp để đưa từ dạng thông tin đã được mã hóa về dạng thông tin ban đầu, quá trình ngược của mã hóa.[[2]](https://vi.m.wikipedia.org/wiki/M%C3%A3_h%C3%B3a)

#### Các thuật toán mã hoá cơ bản:

* Mã hoá cổ điển:
  + Mã hoá cổ điển đã có lịch sử hàng thế kỷ và được áp dụng nhiều cho chiến tranh. Mã hóa cổ điển hoạt động dựa trên cơ sở bảng chữ cái nhằm biến chữ cái này thành chữ cái khác. Mã hoá cổ điển chủ yếu được thực hiện bằng tay hoặc các máy móc đơn giản, ngay cả cỗ máy Enigma cũng được xếp vào mã hoá cổ điển dù dùng những công nghệ hiện đại nhất trong thời kỳ Thế chiến II
  + Phương thức mã hoá cổ điển chủ yếu dựa trên mật mã hoán vị và mật mã thay thế. Mật mã thay thế là việc các kí tự này được thay đổi thành các ký tự khác theo một quy tắc nhất định nhằm phá vỡ ý nghĩa của câu gốc. Ví dụ như sử dụng mã hóa Ceasar ta có thể biến đổi từ "Hello" với độ dịch là 3 trở thành "Vrjjn". Mật mã hoán vị được hoạt động theo cách giữ nguyên các kí tự của văn bản gốc nhưng đổi chỗ các kí tự theo một trật tự nhất định. Các phương thức hiện đại hơn như DES hay AES thực hiện việc sử dụng đồng thời cả hai phương thức hoán vị và thay thế một số lần
  + Tuy nhiên mã hoá cổ điển còn mang trong mình nhiều nhược điểm như kém an toàn, chỉ cần bị tiết lộ thuật toán để mã hoá hoặc sử dụng phân tích tần suất để thử thuật toán thì thông tin coi như là đã không còn an toàn. Thứ hai là việc mã hóa cổ điển chỉ mã hoá các kí tự chữ nên không có nhiều ý nghĩa hiện nay khi còn người hoạt động trên môi trường mạng, nơi mà các phương thức mã hoá hiện đại thường mã hóa các bit dữ liệu.
* Mã hoá một chiều(Hash)[[3]](https://codelearn.io/sharing/hash-la-gi-va-hash-dung-de-lam-gi)
  + Hashing là một thuật toán mã hoá được đánh giá cao và có nhiều ứng dụng trong việc bảo mật thông tin người dùng - những thông tin ta không cần phải dịch lại nguyên gốc. Hash (Hàm băm) là quá trình biến đổi đầu vào thành một đầu ra có độ dài được quy định sẵn sử dụng nhiều thuật toán và các công thức toán học khác nhau. Cùng một giá trị đầu vào thì sau khi băm nó sẽ cùng cho ra một đầu ra là như nhau
  + Hiện nay phổ biến nhất là các hàm băm như MD5, SHA-1, SHA-2
  + Mã hoá một chiều có rất nhiều ứng dụng hiện nay. Chúng được sử dụng để kiểm tra tính toàn vẹn của thông tin bằng cách so sánh các giá trị băm lúc gửi và nhận với nhau nếu chúng không trùng khớp tức là tệp tin đã có sự can thiệp bởi bên thứ ba. Một ứng dụng phổ biến nữa là kiểm tra mật khẩu nhập vào. Mật khẩu sau khi nhập vào sẽ được chạy qua hàm băm rồi sau đó kiểm tra lại với chuỗi băm được lưu trong cơ sở dữ liệu để chứng thực. Bằng cách này ngay cả khi kẻ gian có xâm nhập được vào cơ sở dữ liệu thì chúng cũng chỉ thấy được hàm băm mà không thấy được mật khẩu. Để an toàn hơn các hệ thống lớn đã thêm vào các giá trị muối (salt) vào mật khẩu gốc, do đó nếu bị tấn công thì kẻ xấu chưa thể có được mật khẩu do đã được thêm các giá trị muối

[](https://commons.wikimedia.org/wiki/File:CPT-Hashing-Password-Hashed.svg)

Hình 2.2 Kiểm tra mật khẩu bằng hàm băm

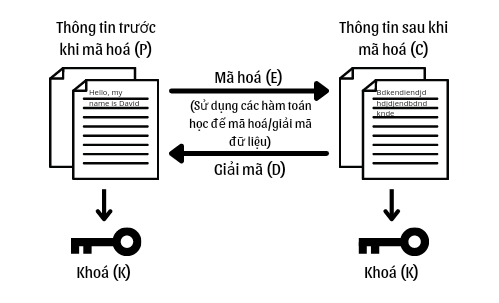
* Mã hoá đối xứng (symmetric key encryption)
* Mã hoá bất đối xứng (public key encryption)
* Ứng dụng của việc mã hoá trong đời sống hằng ngày:

Mã hoá có vai trò vô cùng quan trọng trong đời sống hiện nay, đặc biệt trong môi trường Internet. Nó là công cụ để bảo vệ dữ liệu và cũng như là nền tảng cho nhiều công nghệ khác

* + Đảm bảo tính bí mật, toàn vẹn của thông tin khi thông tin được đẩy lên mạng
  + Đảm bảo tính toàn vẹn của tài liệu ta tải xuống từ Internet
  + Truyền thông tin một cách bí mật
  + Đặc biệt trong các lĩnh vực ngân hàng, các loại tiền mã hoá
  + Là nền tảng của các công nghệ như kĩ thuật chữ ký điện tử, hệ thống PIK, …

### **Hệ thống mã hoá**

* Hệ thống mã hoá là một tập hợp các quy tắc để mã hoá thông tin
* Các thành tố cơ bản của hệ thống mã hoá:[1]
* Thông tin trước khi mã hóa, ký hiệu là P (Plaintext).
* Thông tin sau khi mã hóa, ký hiệu là C (Ciphertext).
* Chìa khóa, ký hiệu là K (Key).
* Phương pháp mã hóa/giải mã, ký hiệu là E/D (Encryption/Decryption).
* Quy trình mã hoá, giải mã
* Quá trình mã hóa được tiến hành bằng cách áp dụng hàm toán học E lên thông tin P, vốn được biểu diễn dưới dạng số, để trở thành thông tin đã mã hóa C
* Quá trình giải mã được tiến hành ngược lại: áp dụng hàm D lên thông tin C để được thông tin đã giải mã P



Hình 2.3 Quy trình mã hoá cơ bản

## Các hệ thống mã hoá cơ bản

### **Hệ thống mã hoá ASCII**

#### Tổng quan về ASCII:

ASCII (tên đầy đủ: American Standard Code for Information Interchange - Chuẩn mã trao đổi thông tin Hoa Kỳ)

#### Lịch sử ra đời và phát triển của ASCII [4]

* 1693. Hiệp hội tiêu chuẩn Hoa Kỳ (American Standards Association, ASA) công bố và trở thành tiêu chuẩn
* 1964. IBM cố gắng tạo phát triển OS/360 để chuyển đổi giữa ASCII và EBCDIC (phần mở rộng của mã punch-card cũ). Nhưng những nỗ lực này đã thất bại
* 1964 – 1981. Trong giai đoạn 18 năm, ASCII gặp rất nhiều khó khăn . Các công ty ở Mỹ cho rằng ASCII là không cần thiết
* Năm 1981. Các máy tính IBM đầu tiên được sinh ra và đánh dấu sự bắt đầu khi ASCII đã thực sự được sử dụng trong máy tính. Đó là một mốc quan trọng cho ASCII thực sự.. Kể từ đó mỗi máy tính đã được thực hiện là ASCII tương thích càng nhiều càng tốt.. tiêu chuẩn này được tạo ra như CBM ASCII và những người khác cụ thể cho các máy nhất định nhưng nhìn chung ASCII sắp được lan truyền rộng rãi trên toàn cầu.
* Hiện nay ASCII vẫn còn được sử dụng ở một số hệ điều hành tuy nhiên không còn được phổ biến do sự phát triển của một bảng mã mới – bảng mã Unicode

#### Các kí tự trong bảng mã ASCII tiêu chuẩn:

* Các ký tự điều khiển

Các ký tự từ 0 đến 31 và ký tự 127 (theo hệ thập phân) là các ký tự điều khiển tức là nó chỉ hiển thị trong môi trường DOS, và một vài kí tự sẽ thực hiện lệnh. Chẳng hạn như: ký tự 7 - BEL nghĩa là tiếng kêu (bạn sẽ nghe thấy tiếng bip), ký tự 10 - LF nghĩa là xuống dòng, …

[](https://www.sites.google.com/site/ttkuzze/hoctap/c-c/bang-ma-ascii)

Hình 2.4 Bảng mã ASCII - Các kí tự điều khiển

* Các kí tự có thể hiển thị được cho người dùng
* Các dấu câu và dấu kết thúc câu (“!”, “?”, “.”, “,”, “:”, “;”,…)
* Dấu cách (" ")
* . Các ký tự đặc biệt và các phép tính (“@”, “#”, “$”, “%”, “^”, “&”, “\*”, “/”, “+”, “-”, “\_”, “]”, “}”, “~”,…)
* Các ký tự chữ thường (“a”, “b”, “c”, “d”, “e”, “f”, “g”, “h”, “i”, “j”, “k”, “l”, “m”, “n”, “o”, “p”, “q”,…)
* Các ký tự chữ in hoa (“A”, “B”, “C”, “D”, “E”, “F”, “G”, “H”, “I”, “J”, “K”, “L”, “M”,…).

#### Bảng mã ASCII mở rộng

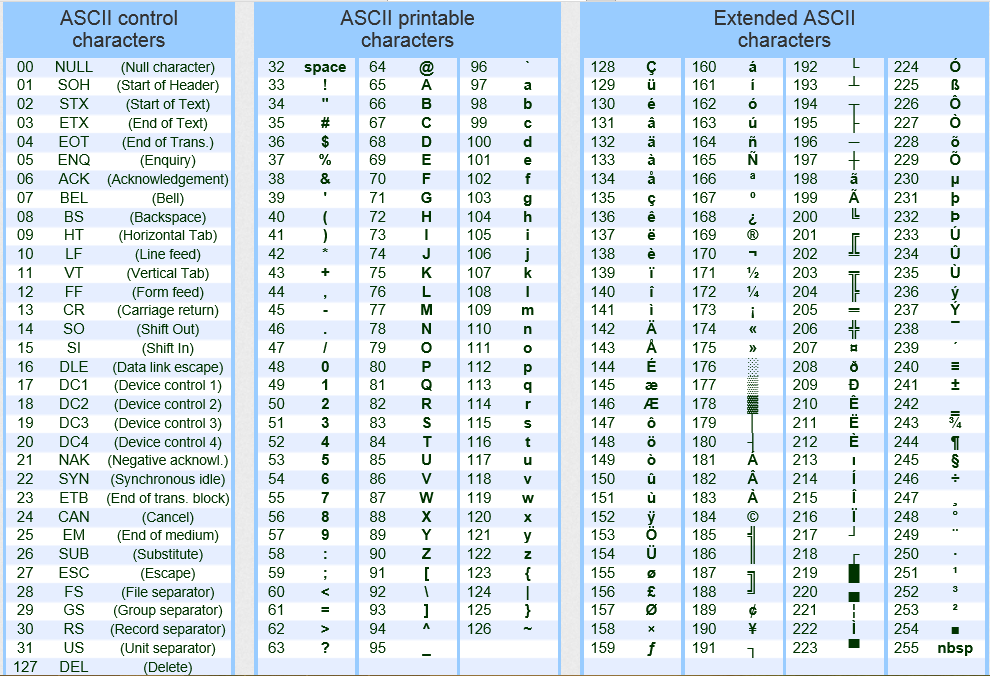
Trên thực tế, mỗi quốc gia lại sử dụng một ngôn ngữ khác nhau với những ký tự khác nhau. Bảng mã ASCII mở rộng đã được ra đời nhằm đáp ứng sự đa dạng trong ngôn ngữ ấy.

Nếu như bảng mã ASCII cơ bản sử dụng 7 bit để biểu thị các ký tự thì bảng ASCII mở rộng sử dụng 8 bit. Vì vậy, bảng này còn được gọi với tên khác là bảng mã ASCII 8 bit.

Sự ra đời của bảng ASCII mở rộng chính là một thành công vang dội của ngành công nghệ thông tin. Kể từ đây, máy tính đã có thể đọc và hiển thị được rất nhiều ngôn ngữ khác nhau, phù hợp với văn hóa của từng quốc gia.

Cùng với sự ra đời của bảng mã ASCII mở rộng, thị trường máy tính và công nghệ thông tin cũng chứng kiến một bước phát triển nhảy vọt. Máy tính dần trở nên phổ biến hơn tại các quốc gia, công nghệ thông tin được áp dụng nhiều hơn vào đời sống.

Bảng mở rộng có số lượng ký tự gấp đôi bảng tiêu chuẩn (256 ký tự). Trong đó bao gồm 128 ký tự của bảng chuẩn, còn lại là các chữ có dấu, các phép toán và các ký tự trang trí. Tiếng Việt của chúng ta cũng được hiển thị trên máy tính là nhờ có bảng ASCII mở rộng này.



Hình 2.5 Bảng mã ASCII mở rộng

### **Hệ thống mã hóa Unicode**

Sự ra đời của bảng mã ASCII là một bước đột phá trong việc đưa ra chuẩn chung để mã hoá dữ liệu, giúp các máy tính trên toàn cầu có thể dễ dàng giao tiếp với nhau và giao tiếp với con người. Tuy nhiên việc bảng mã ASCII tiêu chuẩn chỉ mã hóa 128 kí tự, khi bảng mã ASCII được mở rộng sử dụng 8 bit để mã hoá một kí tự thì nó cũng chỉ biểu diễn 255 kí tự khác nhau. Con số này xét ở thời điểm hiện tại là quá nhỏ khi mỗi vùng đều có một hệ thống bảng chữ cái khác nhau, ví dụ như bảng chữ cái của tiếng Trung, tiếng Nhật hay Tiếng Hàn sử dụng chữ tượng hình, hay bảng chữ cái tiếng Việt có các dấu câu. Vậy nên bảng mã Unicode được ra đời nhằm giải quyết vấn đề đó bằng cách dùng nhiều bit hơn để mã hóa một kí tự

#### Sơ lược về bảng mã Unicode

Unicode (hay gọi là mã thống nhất; mã đơn nhất) 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, kể cả các ngôn ngữ sử dụng ký tự tượng hình phức tạp như chữ Hán của tiếng Trung Quốc, tiếng Nhật, chữ Nôm của tiếng Việt, v.v. Phiên bản đầu tiên của Unicode được Hiệp hội Unicode xuất bản lần đầu tiên vào năm 1991. Nó còn có một tên gọi khác là Universal Coded Character Set (Bộ ký tự được mã hóa chung - UCS). Chuẩn Unicode không chỉ quy định bộ mã, mà còn cả cách dựng hình, cách mã hóa (sử dụng 1, 2, 3 hay 4 byte để biểu diễn một ký tự)

#### Kho chữ của hệ thống mã hóa Unicode

Unicode chiếm trước 1.114.112 (= 220+216) mã chữ, và hiện nay đã gán ký hiệu cho hơn 96000 mã chữ. 256 mã đầu tiên phù hợp với ISO 8859-1, là cách mã hóa ký tự phổ biến nhất trong "thế giới phương Tây"; do đó, 128 ký tự đầu tiên còn được định danh theo ASCII.

Không gian mã Unicode cho các ký tự được chia thành 17 mặt phẳng (plane) và mỗi mặt phẳng có 65536 code point. Mặt phẳng đầu tiên (plane 0), "Mặt phẳng đa ngôn ngữ căn bản" (Basic Multilingual Plane - BMP), là nơi mà đa số các ký hiệu được gán mã. BMP chứa các ký hiệu cho hầu hết các ngôn ngữ hiện đại, và một số lượng lớn các ký tự đặc biệt. Đa số các code point được phân bố trong BMP được dùng để mã hóa các ngôn ngữ CJKV (Hán-Nhật-Hàn-Việt).

Hai mặt phẳng tiếp theo được dùng cho các ký tự "đồ họa". Mặt phẳng 1, "Mặt phẳng đa ngôn ngữ bổ sung" (Supplementary Multilingual Plane - SMP), được dùng chủ yếu cho các loại chữ viết cổ, ví dụ Egyptian hieroglyph (chưa được mã hóa), nhưng cũng còn được dùng cho các ký hiệu âm nhạc. Mặt phẳng 2, (Supplementary Ideographic Plane - SIP), được dùng cho khoảng 40000 chữ Trung Quốc ít gặp mà đa số là các ký hiệu cổ, ngoài ra cũng có một số ký hiệu hiện đại. Mặt phẳng 14 hiện chứa một số các ký tự thẻ ngôn ngữ không được khuyến khích và một số ký hiệu lựa chọn biến thể. Mặt phẳng 15 và Mặt phẳng 16 được mở cho các sử dụng cá nhân.

Vẫn còn nhiều tranh luận giữa các chuyên gia về ngôn ngữ CJK (Hoa-Nhật-Hàn), đặc biệt là các chuyên gia người Nhật, về nhu cầu và lợi ích kỹ thuật của việc "thống nhất chữ Hoa", tức là việc chuyển những bộ chữ Hoa và chữ Nhật vào trong một bộ chữ hợp nhất. (Xem thêm mã hóa chữ Hoa)

Kho ≈220 điểm mã bảo đảm sự tương thích với bộ mã UTF-16. Việc mới chỉ dùng hết có 10% kho chữ cho thấy rằng kho chữ cỡ ≈20 bit này khó bị đầy trong một tương lai gần.[[5]](https://vi.m.wikipedia.org/wiki/Unicode)

#### Các bảng mã của Unicode

Unicode không chỉ quy định bộ mã mà còn quy định cách mã hoá dữ liệu. Từ đó có nhiều định dạng mã hoá được sinh ra, nhưng phổ biến nhất là ba định dạng UTF-8, UTF-16, UTF-32.

* Định dạng mã hoá UTF-8: Mỗi ký tự sẽ được mã hóa thành chuỗi 1,2,3 hay 4 bit để tương thích với bộ mã ASCII
* Định dạng mã hoá UTF-16: Mỗi ký tự sẽ được mã hóa bởi một hay hai chuỗi 16 bit
* Định dạng mã hoá UTF-32: Mỗi kí tự được biểu diễn trong một chuỗi cố định 32 bit

#### Cách bảng mã Unicode mã hoá hiển thị tiếng Việt

Tiếng Việt được xếp vào họ La-tinh mở rộng 1 (Latinh Extended 1), tuy rằng các ký tự Việt phân bố không tập trung, nhưng có 1 thuận lợi rất lớn là do tiếng Việt được xếp vào họ La-tinh nên đây là thành phần cơ bản của hầu hết tất cả các font chữ Unicode, có nghĩa là tiếng Việt có mặt ở mọi nơi trong bất kỳ của ngôn ngữ nào, và như thế có thể đọc được tiếng Việt ở mọi nơi có cài đặt font Unicode. Trong khi đó các ngôn ngữ không thuộc họ la-tinh như Trung Quốc, Nhật Bản, Lào, Thái Lan,... thì không phải lúc nào cũng có sẵn trong các font Unicode.

Tiếng Việt trong Unicode có thể có 2 dạng: ký tự dựng sẵn và ký tự tổ hợp. Unicode có đủ 134 ký tự cho tất cả chữ hoa và chữ thường trong bảng chữ cái tiếng Việt, đồng thời có mã cho 5 dấu thanh (Huyền, sắc, hỏi, ngã, nặng) để tạo ra các ký tự Việt dạng tổ hợp, ngoài ra Unicode còn có dấu riêng để biểu diễn đơn vị tiền đồng Việt nam(₫).[6]

Khi soạn thảo tiếng Việt ta cần một bộ gõ có cài bảng mã có hỗ trợ tiếng Việt, tiêu biểu là Unicode, VNI, TCVN3. Tuy nhiên có nhược điểm với phông chữ VNI và TCVN3 là khi chuyển file đến một máy tính không có font phù hợp thì tài liệu không thể đọc được. Các văn bản hiện nay thường sử dụng bộ mã Unicode để khắc phục hạn chế này.

# **KẾT LUẬN**

Có thể thấy rằng, trong thế giới hiện đại, hệ nhị phân đã trở thành ngôn ngữ của ngành khoa học điện tử và máy tính. Nó cũng là cơ sở cho mã nhị phân được sử dụng để soạn thảo dữ liệu trên máy tính. Bên cạnh đó, thông tin là một khái niệm trừu tượng mà máy tính không xử lí trực tiếp được, việc mã hóa thông tin giúp chuyển đổi thành các kí hiệu máy có thể hiểu được và xử lí.

# **TÀI LIỆU THAM KHẢO**

1. Các tài liệu, sách báo, tạp chí

[4] Đinh Huyền Trang - Lịch sử phát triển và ra đời của bảng mã ASCII

[6] Đặng Minh Tuấn - Tiếng Việt và bộ mã Unicode

1. Các website

[1] Wikipedia - Máy Enigma

<https://vi.m.wikipedia.org/wiki/M%C3%A1y_Enigma>

Ngày truy cập: 28/10/2022

[2] Wikipedia - Mã hoá

<https://vi.m.wikipedia.org/wiki/M%C3%A3_h%C3%B3a>

Ngày truy cập: 28/10/1022

[3] Codelearn.io - Hash Là Gì Và Hash Dùng Để Làm Gì?

<https://codelearn.io/sharing/hash-la-gi-va-hash-dung-de-lam-gi>

Ngày truy cập: 29/10/2022

[5] Wikipedia - Unicode

<https://vi.m.wikipedia.org/wiki/Unicode>

Ngày truy cập: 29/10/2022