

Informatique 2C

A. Ridard

Représentation de l'information



Table des matières

I.	Introduction	4
II.	Les entiers (naturels)	4
III.	Les caractères (ASCII)	6
IV.	Les images (matricielles)	7

I. Introduction

Dans ce chapitre on s'intéresse à la manière dont un ordinateur représente l'information afin de pouvoir la traiter automatiquement.



I Le mot informatique est la concaténation de « information » et « automatique ».

Dans un ordinateur, toutes les informations ^[1] sont représentées à l'aide de deux chiffres 0 et 1, appelés chiffres binaires (binary digits) ou plus simplement *bits*.

Dans la mémoire d'un ordinateur, ces chiffres binaires sont regroupés en *octets* (bytes) c'est à dire en « paquets » de 8, puis organisés en *mots machine* (words) de 2, 4 ou 8 octets (pour les machines les plus courantes aujourd'hui, dites de 64 bits).

Ce regroupement des bits en octets ou mots machine permet de représenter (et manipuler) d'autres données que des 0 ou des 1, comme par exemple des nombres (entiers et approximation de réels), des caractères, mais aussi des images, des sons et des vidéos!

Pour cela, il est nécessaire de disposer d'encodages pour représenter ces informations.

II. Les entiers (naturels)

L'encodage le plus simple est celui des nombres entiers naturels.

Il consiste à *interpréter* un octet ou un mot machine comme un entier écrit en base 2.

Rappel : écriture en base 10

Un entier naturel en base 10 est une séquence de chiffres entre 0 et 9.

Chaque chiffre est associé à une puissance de 10 selon sa position ^[2].

séquence	6	1	0	2	7
position	4	3	2	1	0
poids	10^4	10^3	10^2	10^1	10^0

La valeur de la séquence 61027 est l'entier n calculé de la manière suivante :

$$n = 6 \cdot 10^4 + 1 \cdot 10^3 + 0 \cdot 10^2 + 2 \cdot 10^1 + 7 \cdot 10^0$$

Le chiffre le plus à droite est dit de poids faible, et celui le plus à gauche de poids fort.

Écriture en base 2

Un entier naturel en base 2 est une séquence de chiffres binaires (0 ou 1).

Chaque chiffre est associé à une puissance de 2 selon sa position ^[3].

séquence	0	1	0	0	1	1	0	1
position	7	6	5	4	3	2	1	0
poids	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0

La valeur de la séquence 01001101 est l'entier n calculé de la manière suivante :

$$n = 0 \cdot 2^7 + 1 \cdot 2^6 + 0 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 77$$

Le bit le plus à droite est dit de poids faible, et celui le plus à gauche de poids fort.

[1]. Données mais aussi programmes

[2]. De la droite vers la gauche, en commençant par la position 0

[3]. De la droite vers la gauche, en commençant par la position 0



Donner la valeur (décimale) des séquences binaires suivantes :

1. 10101101
2. 01110010
3. 1111
4. 11111111

Le passage de la base 10 à la base 2 n'est pas aussi immédiat.

On peut effectuer des divisions euclidiennes successives ou bien utiliser le tableau de conversion suivant et la méthode décrite ci-dessous.

poids	...	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
valeur	...	128	64	32	16	8	4	2	1
bit	...								



Méthode pour compléter le tableau de conversion

1. Déterminer le plus grand poids inférieur ou égal à l'entier à convertir
2. Le bit associé à ce poids est 1
3. Soustraire ce poids à l'entier à convertir
4. Recommencer avec ce nouvel entier tant qu'il est différent de 0
5. En complétant les cases vides ^a par des 0, on obtient le résultat

^a. A droite du bit de poids fort



Écrire en base 2 les entiers suivants :

1. 14
2. 218
3. 42
4. 57

Écriture en base 16 (hexadécimale)

Un entier naturel en base 16 est une séquence de chiffres hexadécimaux :

$$0, 1, \dots, 9, A, B, \dots, F$$

La valeur de chaque lettre est donnée par le tableau suivant :

lettre	A	B	C	D	E	F
valeur	10	11	12	13	14	15

Chaque chiffre est associé à une puissance de 16 selon sa position ^[4].

séquence	2	A	4	D
position	3	2	1	0
poids	16^3	16^2	16^1	16^0

[4]. De la droite vers la gauche, en commençant par la position 0

La valeur de la séquence 2A4D est l'entier n calculé de la manière suivante :

$$n = 2 \cdot 16^3 + 10 \cdot 16^2 + 4 \cdot 16^1 + 13 \cdot 16^0 = 10829$$

La base 16 est souvent utilisée pour simplifier l'écriture en base 2. En effet, on passe facilement de la base 2 à la base 16 en regroupant les chiffres binaires par 4 :

$$\begin{array}{cccc} A & 5 & F & 3 \\ \hline 1010 & 0101 & 1111 & 0011 \end{array}$$

Inversement, pour passer de la base 16 à la base 2, il suffit d'écrire chaque chiffre hexadécimal en binaire avec 4 bits.



On pourra utiliser la notation indicielle pour préciser la base utilisée :

- $101_2 = 5$
- $101_{16} = 257$



Donner la valeur (décimale) des séquences hexadécimales suivantes :

1. 31
2. 4D



Transformer en binaire les séquences hexadécimales suivantes :

1. 15
2. 6DF

III. Les caractères (ASCII)

La représentation des caractères dans un ordinateur est l'élément clé pour stocker ou échanger des textes.

En théorie, c'est très simple, il suffit d'associer un numéro unique à chaque caractère, mais en pratique, l'encodage doit être :

- le même pour tous les ordinateurs qui communiquent entre eux
- capable de représenter tous les caractères ^[5]
- compact pour économiser la mémoire ou le volume des échanges réseaux

Dans les années 50, les différents encodages de caractères, présents dans les ordinateurs ou les imprimantes, étaient incompatibles entre eux et imposaient l'utilisation de nombreux programmes de conversion.

Au début des années 60, l'ANSI ^[6] propose une norme de codage des caractères appelée ASCII ^[7]. Elle définit un jeu de 128 caractères, chacun étant représenté par un octet.



Même si 7 bits sont suffisants pour représenter 128 caractères ^a, en pratique chaque caractère occupe 1 octet (8 bits) en mémoire.

En effet, le bit de poids fort est utilisé pour une « somme de contrôle » afin de détecter d'éventuelles erreurs de transmission, mais ce « bit de parité » sera ignoré dans ce cours.

^a. De 000 0000 à 111 1111

[5]. Y compris les caractères « non imprimables » qui correspondent à des actions (passer à la ligne, ...), mais aussi à des commandes de protocoles de communication (début de texte, ...)

[6]. American National Standards Institute

[7]. American Standard Code for Information Interchange

Cette correspondance est résumée dans la table ASCII :

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

Pour trouver l'octet ^[8] correspondant à un caractère, il suffit de concaténer le chiffre hexadécimal de sa ligne avec celui de sa colonne.

Par exemple :

- le caractère A correspond au nombre 41_{16} c'est à dire à l'octet **0100 0001**
- le caractère + correspond au nombre $2B_{16}$ c'est à dire à l'octet **0010 1011**

La table ASCII contient plusieurs catégories de caractères :

- les lettres de l'alphabet latin en majuscule et en minuscule
- les chiffres de 0 à 9
- des signes de ponctuations, des parenthèses ou des crochets
- des opérateurs arithmétiques
- des caractères spéciaux ^[9] (entre 00 et 20)

Un texte codé en ASCII est alors simplement la suite d'octets correspondant à cette chaînes de caractères.



Donner le codage ASCII, en hexadécimal, de "Ceci est un texte!".



Donner le codage ASCII, en binaire, de "Gymnase".

IV. Les images (matricielles)

Les images **vectérielles** sont des images numériques construites à partir de primitives géométriques (segments de droites, arcs de cercles, polygones, ...) auxquelles on peut appliquer différentes transformations (translations, symétries, rotations, ...).

Créées à partir d'équations mathématiques, ces images vectorielles sont stockées dans des fichiers très légers (.ps et .eps) et peuvent être redimensionnées sans perdre en qualité. Malheureusement, elles sont limitées en termes de réalisme et donc inutilisables en photographie par exemple.

A l'inverse, les images **matricielles** sont représentées à l'aide de tableaux ^[10] de pixels.

IMAGE VECTORIELLE

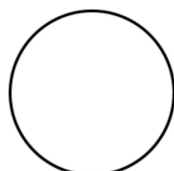
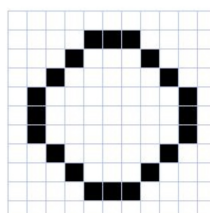


IMAGE MATRICIELLE



[8]. Il est représenté par un nombre hexadécimal à deux chiffres

[9]. Des caractères « blancs » (espace, tabulation, ...), la suppression, l'effacement, ..., mais aussi des caractères « de contrôle » utiles pour les protocoles de communication, le contrôle de périphériques, ...

[10]. Ces tableaux à deux dimensions (lignes et colonnes) sont appelés matrices en Mathématiques

Cette section fait l'objet d'un jupyter notebook disponible sur [moodle](#).

Ce cours s'appuie sur :

- le site internet [Modulo](#) développé par les concepteurs du programme
- l'ouvrage [Numérique et sciences informatiques](#) aux éditions Ellipses