

Représentation numérique de l'information

Où tout se réduit à des 0 et à des 1.

Jean-Marc.Vincent@imag.fr



Novembre 2016

REPRÉSENTATION DE L'INFORMATION

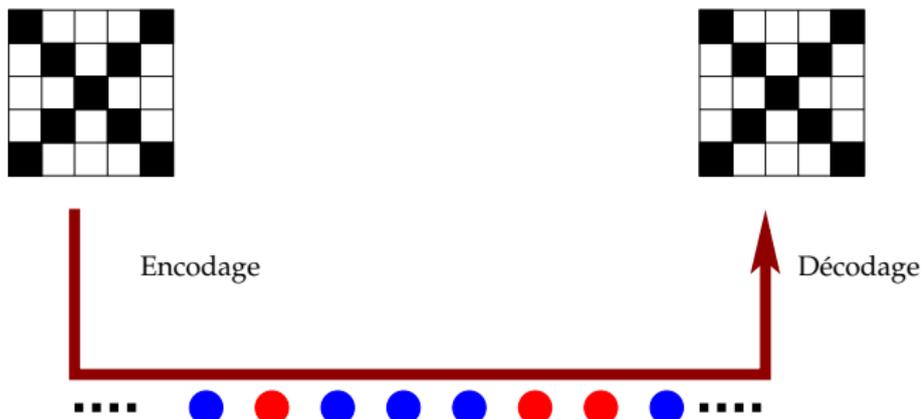
1 **LE PROBLÈME : représentation numérique l'information**

2 CODAGE DES NOMBRES

3 CODAGE DE TEXTES

4 BIBLIOGRAPHIE

INTRODUCTION AU PROBLÈME : TRANSMISSION D'UNE IMAGE SIMPLE



Démarche :

- 1 représentation numérique d'une image (codage commun, ici pixels noir/blanc)
- 2 protocole de communication (spécification)
 - algorithme de codage
 - algorithme de décodage
- 3 test sur des exemples
- 4 Preuve

SYNTHÈSE DE L'ACTIVITÉ DÉBRANCHÉE

UN MONDE NUMÉRIQUE

Numérique

- ▶ **Grand Robert 4. Techn.** Se dit de la représentation de données ou de grandeurs physiques sous forme de nombres (par oppos. à analogique) et par ext. de systèmes, dispositifs ou procédés employant ce mode de représentation. | Traitement numérique de l'information ; traitement numérique du son (→ Audionumérique ; D. A. T.), de l'image, du texte (→ Numériser). — Appareil photo, caméra, caméscope numérique. | Signal numérique : signal qui peut prendre en compte un nombre fini de valeurs discrètes. — Transmission numérique : transmission de signaux codés sous la forme de données binaires. | Télévision numérique. | Bouquet* numérique.
N. m. | Le numérique : l'ensemble des techniques de communication qui utilisent des signaux numériques, notamment dans la reproduction des images.
- ▶ **Trésor de la langue française informatisé A. MATH., INFORMAT.** Qui concerne des nombres, qui se présente sous la forme de nombres ou de chiffres, ou qui concerne des opérations sur des nombres. Méthode numérique ; donnée, échelle, évaluation, rapport, valeur numérique ; équation, expression numérique ; coefficient, exposant, notation numérique.

UN MONDE NUMÉRIQUE

Exemples

- ▶ image, son, video
- ▶ médecine
- ▶ automobile
- ▶ agriculture
- ▶ biologie
- ▶ ...

REPRÉSENTATION DE L'INFORMATION

1 LE PROBLÈME : représentation numérique l'information

2 CODAGE DES NOMBRES

3 CODAGE DE TEXTES

4 BIBLIOGRAPHIE

CODAGE DES ENTIERS NATURELS

- ▶ Représentation du nombre (numération additive ou de position)
- ▶ choix d'une base (alphabet = ensemble de symboles)
- ▶ représentation *machine*
 - bit (*binary digit* atome d'information)
 - byte (mot machine élémentaire (de *bite*)) actuellement correspond à 8 bits
 - généralisation :
 - base octale (8),
 - base hexadécimale (16) {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F}
- ▶ autres représentations (*redundant binary representation (RBR)*)

$$n = \sum_{i=0}^k d_i 2^i \text{ avec } d_i \in \{-1, 0, 1\}$$

Attention représentation finies de nombres

LES ENTIERS EN C

Types entiers, en ordre croissant

Type	Taille	Magnitude <i>signed</i> (magnitude <i>minimale</i> exigée par le standard ¹⁰)	Magnitude <i>unsigned</i> (magnitude <i>minimale</i> exigée par le standard ¹⁰)
<code>char</code> , <code>unsigned char</code> , <code>signed char</code> (C89)	≥ 8 bits	-127 à 127	0 à 255 (0xFF en <i>hexadécimal</i> noté avec le préfixe <code>0x</code> de la syntaxe de C)
<code>short</code> (identique à <code>signed short</code>), <code>unsigned short</code>	≥ 16 bits	-32 767 à +32 767	0 à 65 535 (0xFFFF)
<code>int</code> (identique à <code>signed int</code>), <code>unsigned int</code>	≥ 16 bits (taille d'un <i>mot machine</i>)	-32 767 à +32 767	0 à 65 535 (0xFFFF)
<code>long</code> (identique à <code>signed long</code>), <code>unsigned long</code>	≥ 32 bits	-2 147 483 647 à +2 147 483 647	0 à 4 294 967 295 (0xFFFFFFFF)
<code>long long</code> (identique à <code>signed long long</code>), <code>unsigned long long</code> (C99)	≥ 64 bits	-9 223 372 036 854 775 807 à +9 223 372 036 854 775 807	0 à 18 446 744 073 709 551 615 (0xFFFFFFFFFFFFFFFF)

extrait de Wikipedia

LES PUISSANCES DE 2

k	0	1	2	3	4	5	6	7	8	9	10	11	12
2^k	1	2	4	8	16	32	64	128	256	512	1024	2048	4096

$$2^{20} = 1048576$$

...

THE MAGIC TRICK

1	3	5	7	9	11	13	15
17	19	21	23	25	27	29	31
33	35	37	39	41	43	45	47
49	51	53	55	57	59	61	63

2	3	6	7	10	11	14	15
18	19	22	23	26	27	30	31
34	35	38	39	42	43	46	47
50	51	54	55	58	59	62	63

4	5	6	7	12	13	14	15
20	21	22	23	28	29	30	31
36	37	38	39	44	45	46	47
52	53	54	55	60	61	62	63

8	9	10	11	12	13	14	15
24	25	26	27	28	29	30	31
40	41	42	43	44	45	46	47
56	57	58	59	60	61	62	63

16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

PASSAGE D'UNE REPRÉSENTATION À L'AUTRE

Exercices

- 1 Écrire 42 en binaire, en octal, en hexadécimal
- 2 Quel nombre est représenté en binaire par 110011010 ?
- 3 Proposer un algorithme général pour le passage d'une base à une autre. Où sont les difficultés ?
- 4 Écrire un algorithme récursif qui prend en argument un nombre et écrit sa représentation binaire.
- 5 Écrire un algorithme récursif qui prend successivement des caractères 0 ou 1 et qui renvoie le nombre correspondant à cette écriture binaire.

BINAIRE, VOUS AVEZ DIT BINAIRE...

Description de chaque trigramme [\[modifier | modifier le code \]](#)

Trigramme	Sinogramme	Pinyin	unicode	Image naturelle	Qualités	Autres images
	乾	qián	☰ U+2630	le Ciel 天	Créativité, force, initiative	Le créateur , le cheval (bon, vieux, maigre, sauvage), le père, la tête, le rond, le prince, le jade, le métal, le froid le glace, le rouge sombre, un fruit...
	坤	kūn	☷ U+2637	La Terre 地	Disponibilité, adaptabilité, accueil, don de soi	Le réceptif , la vache, la mère, le ventre, une étoffe, un chaudron, l'économie, l'égalité, le veau avec la vache, un grand char, la multitude, le tronc, le sol noir parmi les autres...
	震	zhèn	☳ U+2633	Le Tonnerre 雷	Impulsion, mise en route, secousse	L'éveilleur , le dragon, le fils aîné, le pied, jaune sombre, une grande rue, un roseau ou un jonc...
	巽	xùn	☴ U+2634	Le Vent, le Bois 風	Pénétration, soumission, intériorisation	Le doux , le coq, la fille aînée, les cuisses, le corbeau, le travail, le blanc, le long, le haut, l'indécis...
	離	lí	☲ U+2632	Le Feu 火	Clarté, lucidité, vivacité, éclat	Ce qui s'attache , le faisan, la fille cadette, l'œil, le brillant, la cuirasse et le casque, la lance et les armes, la sécheresse, la tortue, le crabe, l'escargot, l'arbre desséché dans sa partie haute...
	坎	kǎn	☵ U+2635	L'Eau 水	Profondeur, endurance, peur	L'insondable , le porc, le fils cadet, l'oreille, les fosses, les pièges, l'arc et la flèche, le sang, le rouge, la lune, le bois ferme avec beaucoup de marques...
	艮	gèn	☶ U+2636	La Montagne 山	Rigueur, cohésion, calme, solidité	L'immobilisation , le chien, la main, le 3 ^e /la plus jeune fils, le chemin détourné, les pierres, les portes, les fruits, les semences, le bois ferme et noueux...
	兌	duì	☱ U+2631	Le Marais 澤	Aptitude à l'expression et à la communication, joie, légèreté	Le joyeux , le mouton, la 3 ^e /la plus jeune fille, la bouche (& la langue), la magicienne, écraser briser en morceaux, la voisine, le sol dur et dallé...

Yi-Jing *Traité canonique des mutations* (~ -1000 av. JC)



source wikipedia

LES NOMBRES ENTIERS RELATIFS

Idée Utiliser un bit spécifique indiquant le signe, les autres bits codant la valeur absolue du nombre

Étendue du codage avec n bits

$$-(2^{n-1} - 1) \leq N \leq 2^{n-1} - 1.$$

- ▶ 0 est codé deux fois
- ▶ Addition et soustraction moins évidente car l'opération du passage au négatif n'est pas réversible

Bonne idée : Utiliser le complément à 2

- 1 Les nombres positifs sont codés de la même façon qu'en binaire signé
- 2 Un nombre négatif est codé
 - en prenant son opposé,
 - en le représentant en base 2 sur $n-1$ bits,
 - en complétant chaque bit, et
 - en ajoutant 1

Étendue du codage avec n bits

$$-2^{n-1} \leq N \leq 2^{n-1} - 1.$$

EXERCICES

- 1 Coder en complément à 2, sur un octet, des nombres décimaux 24, -24 , 128 et -128 ?
- 2 Pour des entiers relatifs codés sur 8 bits, écrire les nombres 3 et -4 en utilisant le bit de signe puis en utilisant le complément à 2. Faire l'addition des deux nombres dans les 2 formats. Que faut-il remarquer ?
- 3 Coder en complément à 2, sur un octet, des nombres décimaux 24, -24 , 128 et -128 ?
- 4 Sur 4 bits, à quel nombre décimal correspond le nombre codé en complément à 2 par 1111 ?
- 5 Que se passe-t-il si le résultat de l'addition sort de la plage $-2^{n-1} \leq N \leq 2^{n-1} - 1$.

LES NOMBRES RÉELS

La norme IEEE définit la façon de coder un nombre réel sur 32 bits et définit trois composante

- ➊ Signe sur 1 bit
- ➋ Mantisse sur 23 bits
- ➌ Exposant sur 8 bits (biais = $2^7 - 1 = 127$)

La valeur d'un nombre est donnée par :

$$N = (-1)^s \left(1 + \sum_{i=1}^{23} m_i 2^{-i} \right) . 2^e .$$

avec e calculé en supprimant le biais.

Conventions :

- ▶ Le 0 a deux codages différents : 0000...0 ou 1000...0
- ▶ NaN (not a number) est codé 0111...1
- ▶ l'infini $+\infty$: 011111111000...0 et $-\infty$: 111111111000...0

REPRÉSENTATION DE L'INFORMATION

- 1 LE PROBLÈME : représentation numérique l'information
- 2 CODAGE DES NOMBRES
- 3 CODAGE DE TEXTES
- 4 BIBLIOGRAPHIE

CODAGE DE CARACTÈRES

Un codage historique : ASCII (American Standard Coding for Information Interchange)

- ▶ 128 caractères (jeu minimal) : 33 non imprimables, 94 imprimables, et l'espace (considéré comme invisible)
- ▶ Des limitations claires (US-ASCII), liées à son histoire
- ▶ Jusque fin 2007, l'encodage le plus répandu (UTF-8 depuis)
- ▶ 32 premiers caractères de contrôle (pour des machines comme les imprimantes)
 - Caractère 10 : LF (Line Feed), permet de faire avancer le papier dans une imprimante)
 - Caractère 8 : BS (backspace, retour arrière)
- ▶ Pas de description de la structure d'une page (langage de balisage)
- ▶ Beaucoup d'extensions et de propositions nationales pour utiliser le huitième bit et coder les caractères suivants
- ▶ Certaines extensions non compatibles avec le jeu minimal !

TABLE DES CARACTÈRES ASCII

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Source Wikipedia

CODAGE UTF-8

Universal Character Set Transformation Format - 8 bits

- ▶ Permet de représenter les (milliers de) de caractères du répertoire universel (UCS), commun à la norme IS/CEI 10646 et au standard Unicode
- ▶ Compatible avec US-ASCII
- ▶ UTF-8 est approuvé par l'IETF (Internet Engineering Task Force) et le W3C (World Wide Web Consortium)
- ▶ Inventé en 1992 par K. Thompson (Unix)
- ▶ Remarque On parle de points de code (code points) plutôt que de caractères dans UCS

Principe

- 1 Les points de code ayant une valeur (scalaire, en base 10) comprise entre 0 et 127 sont codés sur un seul octet dont le bit de poids fort est nul
- 2 . Les points de code de valeur supérieure à 127 sont codés sur plusieurs octets :
 - les bits de poids forts du premier octet forment une suite de 1 de longueur égale au nombre d'octets utilisés pour coder le caractère
 - les octets suivants ont 10 comme bits de poids forts

REPRÉSENTATION DE L'INFORMATION

- 1 LE PROBLÈME : représentation numérique l'information
- 2 CODAGE DES NOMBRES
- 3 CODAGE DE TEXTES
- 4 **BIBLIOGRAPHIE**

POUR ALLER PLUS LOIN

- ▶ Le site web [Interstice](#) est une mine de renseignements, parfois très détaillés.
Voir en particulier :
 - [Interstice : Tout a un reflet numérique](#)
 - [Nom de code : binaire](#)
- ▶ *Introduction à la science informatique : Pour les enseignants de la discipline en lycée.* Gilles Dowek. et al CRDP Académie de Paris, 2011. Chapitre 1 Représentation numérique de l'information
- ▶ **Introduction aux sciences de l'information**, Jean-Yves Le Boudec, Patrick Thiran et Rüdiger Urbanke, Presses polytechniques et universitaires romandes, 2015
- ▶ *Informatique et sciences du numérique Spécialité ISN en terminale S. Avec des exercices corrigés et des idées de projets.* Gilles Dowek. et al Eyrolles 2012 Chapitres 7-8
- ▶ **Histoires d'algorithmes** Jean-Luc Chabert et al. Belin (2010)