Chapitre 2 : Représentation des données

Enjeu

- Un ordinateur représente chaque donnée par une suite de 0 et de 1 appelés bits.
- On parle de représentation numérique par opposition à une représentation analogique.
- Sous un format numérique on peut :
  - Dupliquer à l'infini les données sans perte de qualité.
  - Compresser les données facilement.

Exemple : diffusion de la musique

- Le disque vinyle (analogique) était difficile à dupliquer.
- On pouvait l'enregistrer sur des cassettes audio (analogique).
- Chaque copie entrainait une perte de qualité.
- Au bout de quelques copies, la qualité était très médiocre !

Exemple : diffusion de la musique

- Le CD audio a permis de mieux diffuser la musique : meilleure qualité, reproduction très simple.
- Le piratage du CD audio a bouleversé l'économie du disque.
- Une fois la musique diffusée sous un format numérique, sa compression et sa diffusion sous la forme de fichier mp3 a rendu possible l'échange de musique sur internet ... et la généralisation du piratage.

La loi

L'article L. 335-4 sanctionne « [...] toute fixation, reproduction, communication ou mise à disposition du public, à titre onéreux ou gratuit, ou toute télédiffusion d'une prestation, d'un phonogramme, d'un vidéogramme ou d'un programme, réalisée sans l'autorisation, lorsqu'elle est exigée, de l'artiste-interprète, du producteur de phonogrammes ou de vidéogrammes ou de l'entreprise de communication audiovisuelle ».

Sanction

L'article L. 335-2 du Code de la propriété sanctionne le délit de contrefaçon d'une peine d’emprisonnement d'une durée pouvant allée jusqu’à trois ans ainsi que d'une amende pouvant atteindre 300 000 euros. Si le délit est commis en bande organisée, les peines sont portées à cinq ans d’emprisonnement et à 500 000 euros d’amende.

Représentation des entiers

- Chiffres romains : pas pratique pour faire des calculs !
- Chiffres arabes : un être humain fait aisément des opérations.

Idem pour un ordinateur

- Certains systèmes de représentation sont plus pratiques que d'autres pour faire des calculs.
- On cherche des systèmes permettant de faire rapidement des calculs notamment des additions grâce à des assemblage de transistors.
- Pour les entiers la base 2 est le plus répandu.

Base 2 sur 8 bits

- On veut représenter 79 en base 2 sur 8 bits
- On va chercher la plus grande puissance de 2 inférieure ou égale à 79.
- $2^0=1$ $2^1=2$ $2^2=4$ $2^3=8$ $2^4=16$ $2^5=32$ $2^6=64$ $2^7=128$
- On trouve 64 !
  79-64=15
- Et on recommence

- 79-64=15
  15-8=7
  7-4=3
  3-2=1
  1-1=0
- 79=64+8+4+2+1
- À partir de la décomposition on trouve la représentation

  0 1 0 0 1 1 1 1
- Chaque bit à un poids
- Le bit de poids (à gauche) fort a comme poids 128.
- Le bit de poids faible a comme poids 1

Exercice

- Représentez en base 2 sur 8 bits 155, 98 et 189

Représentation de 155 en base 2 sur 8 bits

- 155-128=27
  27-16=11
  11-8=3
  3-2=1
  1-1=0

- Représentation : 1001 1011

Représentation de 98 en base 2 sur 8 bits

- 98-64=34
  34-32=2
  2-2=0

- Représentation :
  0110 0010

Représentation de 189 en base 2 sur 8 bits

- 189-128=61
  61-32=29
  29-16=13
  13-8=5
  5-4=1
  1-1=0

- Représentation : 1011 1101

Conversion base 2 sur entier base 10

- Quel est l'entier représenté en base 2 par

  1010 1110
- Il faut faire la somme des bits à 1
- 128+32+8+4+2=174

Exercice

- Quels sont les entiers représentés en base 2 par
  <u>1100 0111</u>
  <u>0101 1010</u>
  <u>1101 1000</u>

Solution

- 1100 0111==> 128+64+4+2+1=199
  0101 1010==> 64+16+8+2=90
  1101 1000==>128+64+16+8=216

Base 2 sur 8 bits

- Le plus petit 0000 0000 ==> 0
- Le plus grand 1111 1111=255
- On peut représenter en base 2 sur 8 bits tous les entiers entre 0 et 255

Base 2 sur N bits

- On peur représenter en base 2 sur N bits les entiers de 0 = 2^{N-1}

Et les entiers signés ?

- La base 2 ne permet de représenter que des entiers positifs ou nuls.
- Un ordinateur a besoin aussi de manipuler des entiers positifs, négatifs ou nuls.
- Pour représenter ces entiers signés on utilisera le complément à 2

Lien avec le langage C

- Type unsigned int ==> base 2
- Type int ==> complément à 2

Complément à 2 sur 8 bits

- On veut représenter X
- Si X< 0

  le premier bit est à 1

  les 7 autres bits représentent 128+X en base 2 sur 7 bits.
- Si X>=0

  le premier bit est à 0

  les 7 autres bits représentent X en base 2 sur 7 bits.

Exemple : représentation de -25 en complément à 2 sur 8 bits

- -25<0 ==> premier bit à 1
- 128-25=103
- 103-64=39
  39-32=7
  7-4=3
  3-2=1
  1-1=0
- Représentation : 1110 0111

Exercice

- Représentez en complément à 2 sur 8 bits -67, 81, -118, -1

Représentation de -67 en complément à 2 sur 8 bits

- -67<0 ==> premier bit à 1
- 128-67=61
- 61=32+16+8+4+1
- Représentation 1011 1101

Représentation de 81 en complément à 2 sur 8 bits

- 81>0 ==> premier bit à 0
- 81=64+16+1
- Représentation 0101 0001

Représentation de -118 en complément à 2 sur 8 bits

- -118<0 ==> premier bit à 1
- 128-118=10
- 10=8+2
- Représentation : 1000 1010

Représentation de -1 en complément à 2 sur 8 bits

- -1<0 ==> premier bit à 1
- 128-1=127
- 127=64+32+16+8+4+2+1
- Représentation : 1111 1111

Conversion complément à 2 vers base 10

- Quel est l'entier représenté en complément à 2 par 1101 1011
- On fait la somme des poids des bits à 1
- Le premier bit (gauche) à comme poids -128
- Les autres bits ont des poids positifs de 1 à 64
- 1101 1011 ==> -128+64+16+8+2+1=-37

Exercice

- Quels sont entiers représentés en complement à 2 par
  1101 1100
  1001 1001
  0111 1110

Solution

- 1101 1100==> -128+64+16+8+4=-36
  1001 1001==> -128+16+8+1=-103
  0111 1110 ==> 64+32+16+8+4+2=126

Complément à 2 sur 8 bits

- Le plus grand : 0111 1111==>127
- Le plus petit 1000 0000 ==> -128
- En complément à 2 sur 8 bits on peutt représenter les entiers de -128 à 127

- Sur N bits on représente les entiers de -2^{N-1} à 2^{N-1} -1

Hexadécimal

- Il ne s'agit pas à proprement parlé d'un système de représentation mais d'une notation qui permet d'écrire de manière plus courte pour un être humain une longue suite de bits
- On part d'une suite de bits
- On la découpe par tranche de 4 bits
- Chaque tranche est représentée par un caractère

Hexadécimal

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

Conversion binaire vers hexadécimal

- 1101 1110 0101 1011

  ==> DE5B

Conversion hexadécimal vers binaire

- AF8B

  ==> 1010 1111 1000 1011

Représentation des réels

- Les représentations des réels sont généralement plus techniques et plus complexes
- Nous allons étudier le format IEEE-754 simple précision sur 32 bits

IEEE-754 simple précision
Etape 1

- On veut représenter X
  - Si X=0 alors X est représenté par 32 bits à 0
  - Sinon on cherche s, e et m vérifiant $X=(-1)^s 2^e (1+m)$

  avec s valant 0 ou 1, e un entier signé et m un réel de l'intervalle [0-1[

- s=signe
  e=exposant
  m=mantisse

Comment trouver s ?

- Si $X<0$, $s=1$
- Si $X>=0$, $s=0$

Comment trouver e et m ?

- 3 cas
  |X|>=2
  |X]<1

  1<=|X|<2

Premier cas |X|>=2

- On divise |X| par 2 jusqu'à obtenir un réel de l'intervalle [1;2[
- e=+ nombre de divisions
- m=dernière valeur -1

Exemple X=-8,8

- X<0 donc s=1
- 8,8/2=4,4

  4,4/2=2,2
  2,2/2=1,1
- e=+3
- m=1,1-1=0,1

Deuxième cas cas |X|<1

- On multiplie |X| par 2 jusqu'à obtenir un réel de l'intervalle [1;2[
- e=- nombre de multiplications
- m=dernière valeur -1

Exemple X=0,325

- X>0 donc s=0
- 0,325x2=0,65

  0,65*2=1,3
- e=-2
- m=1,3-1=0,3

Troisième cas $1<=|X|<2$

- $e=0$
- $m=|X|-1$

Exemple X=-1,4

- s=1
- e=0
- X=1,4-1=0,4

Etape 2 trouver la suite de 32 bits

- 3 champs :
  signe 1 bit exposant 8 bits mantisse 23 bits
- Exposant onn calcule E=127+e et on représente E en base 2 sur 8 bits

Exemple 1 : X=-8,8

- Etape 1 : s=1 e=3 m=0,1
- E=127+e=130 ==> exposant 1000 0010
- 0,1x2=0,2=0+0,2
  0,2x2=0,4=0+0,4 ==> début période

  0,4x2=0,8=0+0,8
  0,8x2=1,6=1+0,6
  0,6x2=1,2=1+0,2 ==> fin période
  0,2x2=0,4=0+0,4
- Mantisse= 0 0011 0011 0011 0011 0011 00
- Représentation finale :
  1 1000 0010 0 0011 0011 0011 0011 0011 00

Etape 3

- On écrit le résultat final en hexadécimal

Exemple 1 : X=-8,8

- Etape 1 : s=1 e=3 m=0,1
- Etape 2 :
  1 1000 0010 0 0011 0011 0011 0011 0011 00
- Etape 3 :
  C10C CCCC

Exemple 2 : X=0,325

- Etape 1 : s=0 e=-2 m=0,3
- E=127+e=125 ==> exposant 0111 1101
- 0,3x2=0,6=0+0,6
  0,6x2=1,2=1+0,2 ==> début période

  0,2x2=0,4=0+0,4

  0,4x2=0,8=0+0,8
  0,8x2=1,6=1+0,6==> fin période
  0,6x2=1,2=1+0,2

- Mantisse= 0 1001 1001 1001 1001 1001 10
- Représentation finale :
  0 0111 1101 0 1001 1001 1001 1001 1001 10

Exemple 2 : X=0,325

- Etape 1 : s=0 e=-2 m=0,3
- Etape 2 :
  0 0111 1101 0 1001 1001 1001 1001 1001 10
- Etape 3 :
  3EA6 6666

Exemple 3 : X=-1,4

- Etape 1 : s=1 e=0 m= 0,4
- E=127+e=127 ==> exposant 0111 1111

0,4x2=0,8=0+0,8 ==> début période
0,8x2=1,6=1+0,6
0,6x2=1,2=1+0,2

0,2x2=0,4=0+0,4 ==> fin période
0,4x2=0,8=0+0,8

- Représentation
1 0111 1111 0110 0110 0110 0110 0110 011

Exemple 3 : X=-1.4

- Etape 1: s=0 e=0 m=0,4
- Etape 2 :

  1 0111 1111 0110 0110 0110 0110 0110 011
- Etape 3 :
  BFB3 3333

EXERCICE

- Représentez dans le format IEEE-754 les réels
  -6.4
  0.2125

  -0.625

  -1.9
  Les résultats seront donnés en hexadécimal

SOLUTIONS

- Représentez dans le format IEEE-754 les réels
  -6.4 = C0CC CCCC
  0.2125=3e59 9999

  -0.625=BF20 0000

  -1.9=BFF3 3333
  Les résultats seront donnés en hexadécimal

Conversion IEEE-754 vers réel

- Voici un réel représenté en IEEE-754
  Quelle est sa valeur ?

  C0F0 0000

- Étape 1 : écriture en binaire
  1100 0000 1111 0000 0000 0000 0000 0000
  s=1
  exposant= 100 000 1
  mantisse=111 0000 0000 0000 0000 0000

Exposant

- Exposant= 1000001
  E= 128+1=129

- e=E-127=2

mantisse

- Mantisse 11100000000000000000000
- Le premier bit de la mantisse a comme poids 2^{-1} le deuxième 2^{-2} puis 2^{-3} et le dernier 2^{-23}
- On fait la somme des poids des bits à 1
- M=1/2+1/4+1/8=7/8

Retrouver X

- $X=(-1)^s 2^e (1+m)$

- $X=(-1)^1 2^2 (1+7/8)=-15/2$
- **Le résultat doit être donné sous la forme d'une fraction irréductible !**

EXERCICE

- Quels sont les réels représentés en IEEE-754 par
  C1C8 0000
  3F60 0000
  C299 0000

SOLUTION

- C1C8 0000= -25
  3F60 0000=7/8
  C299 0000=-76.5

Système de représentation exact

- Si on prend en entier représentable en base 2 (ou en complément à 2), si on calcule sa représentation et on calcule ensuite le réel de départ, on obtient exactement l'entier représentation.
- La base 2 et le complement à 2 sont des systèmes de représentation exacts

Système de représentation approché

- Si on prend un réel et si on le représente en IEEE-754 puis si on cherche la valeur de la représentation, on obtient un réel proche du réel de départ mais qui peut être légèrement différent.
- Le système de représentation est approché

Erreurs de calcul

- Chaque calcul en IEEE-754 est entaché d'une petite erreur.
- Si on fait beaucoup de calculs cette erreur peut s'amplifier et l'incertitude sur le résultat peut devenir dramatique

Représentation des caractères

- 1963 : apparition du code ASCII. Chaque caractère est codé sur 7 bits.

Table ASCII

USASCII code chart

Bits b_7 b_6 b_5 →					0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b_4	b_3	b_2	b_1	Row ↓ \ Column →	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	1	9	HT	EM	)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[	k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	|
1	1	0	1	13	CR	GS	−	=	M	]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

ASCII étendu

- Sur 8 bits
- Caractères accentués ou autres
- Des centaines de standards différents

unicode

- 128 172 caractères
- 32 bits
- Se veut universel
- Très répandu
- Mais des centaines de concurrents

Codage des caractères

- Aucun véritable standard
- De nombreuses normes incompatibles
- De nombreux problèmes pour les développeurs