

Interopérabilité des systèmes d'information

2. Interopérabilité syntaxique

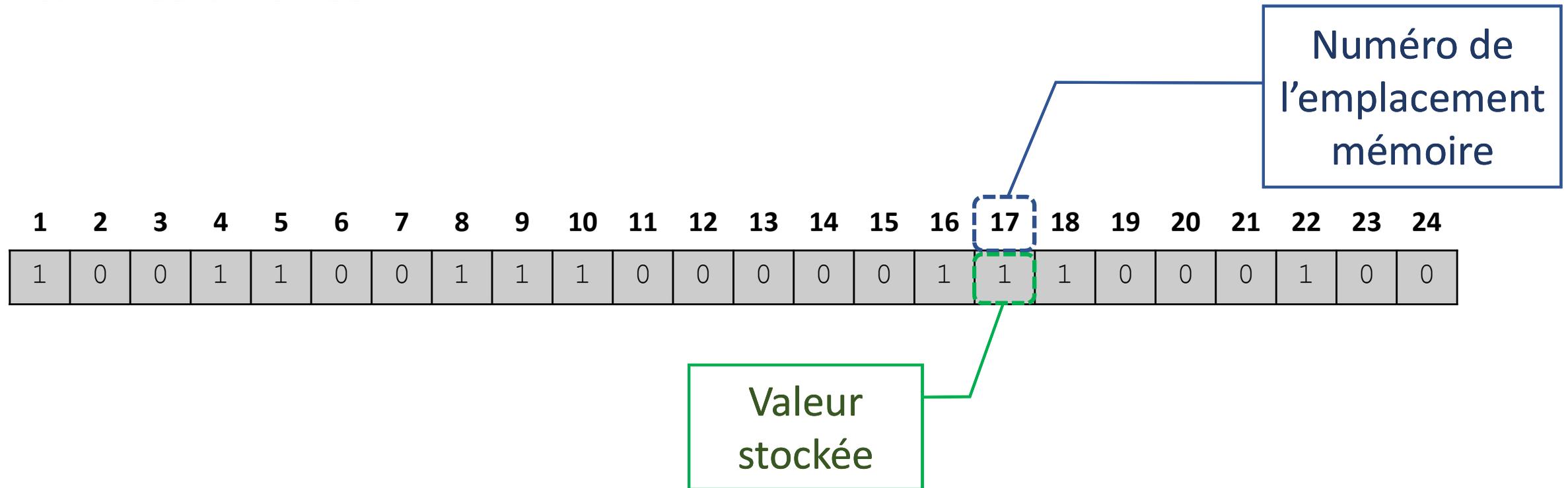
Mastère *Management de la transition industrielle*

Antoine Zimmermann

8 février 2024

Commençons par le commencement

- Les données des systèmes informatiques actuelles sont stockées dans des emplacements qui peuvent être décrits comme une séquence de chiffres binaires :



Comment stocker des nombres ?

- Exercice :
 - Comment stocker les nombres $\{0, 1, 8, 142, 508345\}$ en 8 bits ? En 16 bits ?
 - Comment stocker un nombre négatif comme p.ex. $\{-1, -127\}$?
 - Comment stocker un nombre décimal comme p.ex. $\{3.1415, -1.04, 0.00000123\}$?

Comment stocker des caractères ?

- Tout dans un système informatique doit être stocké comme une suite de bits, y compris les lettres de l'alphabet, la ponctuation, les smileys, les images, les sons, les instructions à exécuter
- Comment stocker en mémoire les lettres {a, b, ..., z, A, B, ..., Z}
- Comment stocker un mot ou chaîne de caractères ? Comment stocker le mot « école » ? La phrase « L'école a 1500 élèves ! »

Rappel : Le protocole Internet (IP)

- IP définit le contenu des paquets permettant de faire « dialoguer » un *client* et un *serveur* situés n'importe où sur le réseau

En-tête IPv4

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Version d'IP				Longueur de l'en-tête				Type de service								Longueur totale															
Identification																Indicateur		Fragment offset													
Durée de vie								Protocole								Somme de contrôle de l'en-tête															
Adresse source																															
Adresse destination																															
Option(s) + remplissage																															

Codage des caractères

- Il existe de nombreuses manières d'encoder des caractères :
https://fr.wikipedia.org/wiki/Codage_des_caract%C3%A8res
- Notamment:
 - ASCII (codage sur 7 bits ou 8 bits)
 - UNICODE (codage sur 8 bits ou 16 bits ou 32 bits)
 - Windows 1252 (8 bits) et variantes

ASCII (7 bits)

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]



- Adopt a Character +
- Emoji +
- Basic Info +
- News
- Events
- Connect +
- Membership +
- Press

 Search...

 U+0D05	 U+00D3	 U+2669	 U+0F0B	 U+201D	 U+05D1	 U+1F600	 U+0CA4
 U+FF4D	 U+FF1B	 U+21A9	 U+0A1D	 U+15E2	 U+FE4F	 U+0CB0	 U+067B
 U+0B67	 U+25D1	<h1>Unicode</h1> <p>Norme informatique pour le codage, la représentation et la manipulation cohérents de textes exprimés dans la plupart des systèmes d'écriture du monde. 144 762 caractères couvrant 159 écritures modernes et historiques, ainsi que des symboles, des emoji et des codes de contrôle et de formatage non visuels.</p>					
 U+1F490	 U+04E9						
 U+0EA7	 U+203A	 U+203B	 U+134C	 U+2019	 U+00D7	 U+2020	 U+0920
 U+1F618	 U+0F4F	 U+0296	 U+1F9E2	 U+1F49A	 U+13EA	 U+1F911	 U+2030

Formats de fichiers

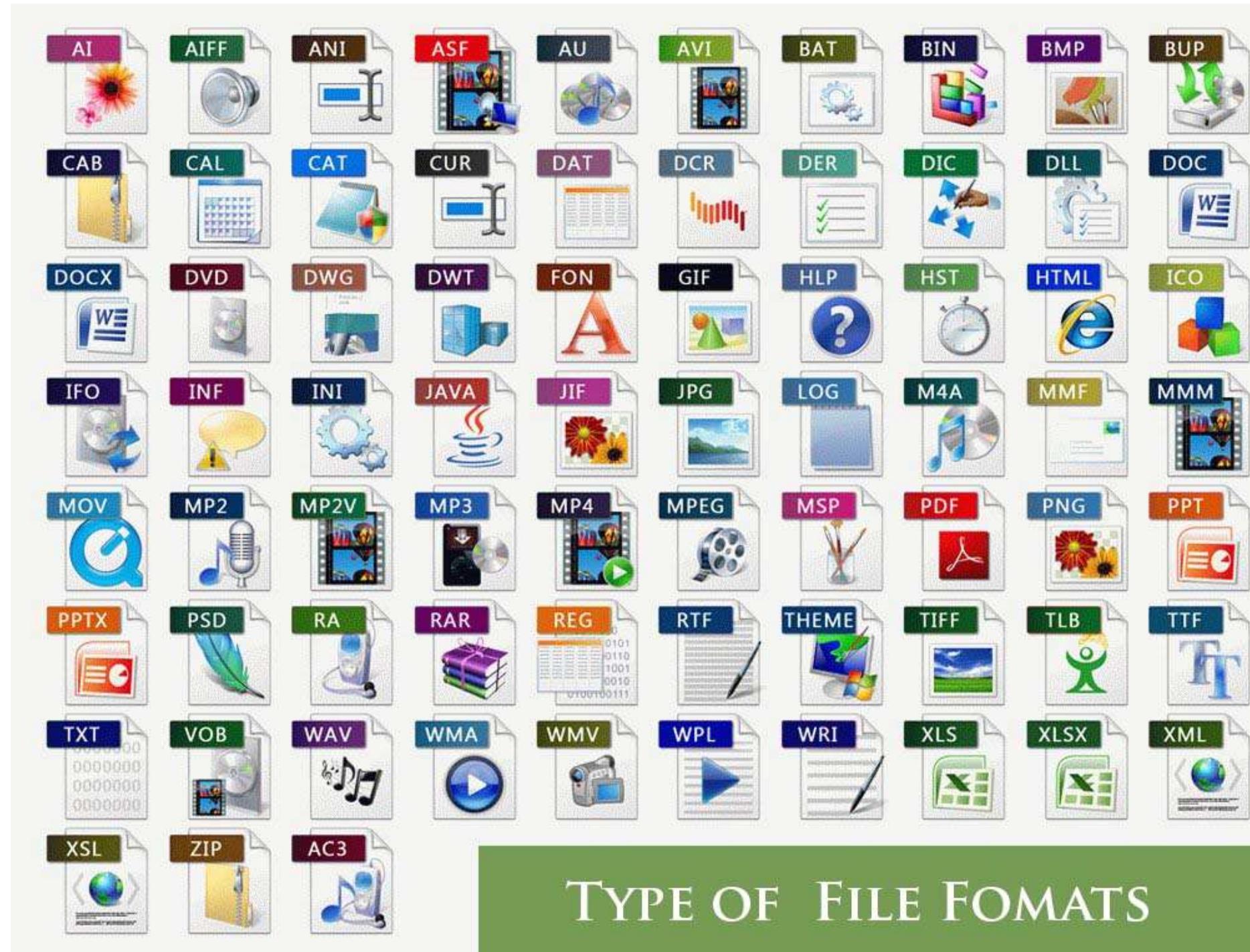
- Un format permet de structurer les données.
- Par exemple : coder un tableau de valeurs avec des lignes et des colonnes ? Préciser que certaines valeurs sont des nombres, d'autres sont des chaînes de caractères, une date, etc.
- Comment représenter un texte mis en forme (gras, italique, taille, couleur) ?
- Comment représenter une image ? Un son ? Un programme exécutable ? Un rapport de panne d'un système de production automatisé, à destination d'un système de réparation intelligent autonome ?

Formats de fichiers

Manière dont l'information est codée pour être transférée ou stockée dans un fichier informatique.

Un format est ce qui permet à une application d'interpréter les données brutes contenues dans un fichier.

Les formats de fichiers peuvent être indiqués dans l'extension du nom du fichier, mais pas forcément.



Encoder un tableau de valeurs (lignes/colonnes)

- Formats légers:
 - CSV: *Comma Separated Values* (valeurs séparées par des virgules)
 - + variantes (séparation par des points-virgules, des tabulations, des espaces)
- Formats lourds:
 - Tableurs (Excel, LibreOffice/OpenOffice calc, Google spreadsheet, etc.)
- Formats d'export de bases de données relationnelles
 - Requêtes SQL d'insertion de données

Encoder des structures hiérarchiques

- De nombreux documents ou jeux de données décrivent une structure hiérarchique:
 - Documents textuels : Livre, parties, chapitres, sections, sous-sections, sous-sous-sections, etc.
 - Description de systèmes complexes : Un élément est composé de sous-éléments eux-mêmes composés de sous-sous-éléments, etc.
- Formats hiérarchiques:
 - XML : *eXtensible Markup Language* (langage de balisage extensible)
 - JSON : *JavaScript Object Notation* (notation objet de JavaScript)

Formats ouverts / formats propriétaires

- Un **format ouvert** est un format dont la *spécification* est accessible publiquement :
 - Permet à tout le monde de concevoir un programme qui décode le format et exploite les données conformes à ce format
- Un **format propriétaire** est conçu par une entreprise qui garde le secret de sa spécification :
 - Il faut utiliser un *logiciel propriétaire* pour pouvoir décoder, éditer et sauvegarder le format correctement
- Avec la démocratisation de l'Internet et du Web, les formats propriétaires sont de plus en plus rares