

Les formats

Débutant – Semaine 4

1

Formats de fichiers

- Qu'est-ce que le codage de données ?
- Qu'est-ce qu'un format de fichier ?

2

Codage ?

123 ↔ Cent vingt trois ↔ CXXIII

Trois écritures de la même valeur à l'aide d'une liste :

- de chiffres (0,1,2, ...)
- de mots (un, deux, dix, cinquante, ...)
- de lettres (I,V,X,L,C, ...)

Codage : Règle définissant l'écriture à l'aide d'une liste donnée

3

En informatique, pourquoi des 0 et des 1 ?

- Une mémoire est un ensemble de bits.
- Un bit n'a que **DEUX états possibles**.
- Ses 2 états sont notés par 0 et 1.

Ainsi on peut dire qu'une mémoire enregistre une suite de 0 et de 1

- Pour enregistrer une information dans une mémoire informatique il suffit donc de définir son codage sous forme de suite de 0 1.

4

Tout est codé en binaire?

Coder du texte ?

- Associer une valeur binaire à chaque lettre
- Par exemple :

A 0 1 0 0 0 0 0 1

B 0 1 0 0 0 0 1 0

Codage dit ASCII : 256 possibilités sur un octet

5

La table ASCII

- À l'origine (7 bits par caractère) : les caractères latins sans accent.
- Puis nombreuses extensions (une par langue) non compatibles entre elles : (ASCII étendu, 8bits par caractère)
- Enfin, une table universelle « utf-8 » (au plus 32 bits pour un caractère ou idéogramme)

盲人摸象 - 各執一端

Traduction : mǎng rén mō xiàng... gè zhí yì duān

Character	Decimal Number	Binary Number	Character	Decimal Number	Binary Number
Blank space	32	0010 0000	^	94	0101 1110
!	33	0010 0001	-	95	0101 1111
"	34	0010 0010	_	96	0110 0000
#	35	0010 0011	`	97	0110 0001
\$	36	0010 0100	~	98	0110 0010
A	65	0100 0001	~	99	0110 0011
B	66	0100 0010	~	100	0110 0100
C	67	0100 0011	~	101	0110 0101
D	68	0100 0100	~	102	0110 0110
E	69	0100 0101	~	103	0110 0111
F	70	0100 0110	~	104	0110 1000
G	71	0100 0111	~	105	0110 1001
H	72	0100 1000	~	106	0110 1010
I	73	0100 1001	~	107	0110 1011
J	74	0100 1010	~	108	0110 1100
K	75	0100 1011	~	109	0110 1101
L	76	0100 1100	~	110	0110 1110
M	77	0100 1101	~	111	0110 1111
N	78	0101 1110	~	112	0111 0000
O	79	0101 1111	~	113	0111 0001
P	80	0101 0000	~	114	0111 0010
Q	81	0101 0001	~	115	0111 0011
R	82	0101 0010	~	116	0111 0100
S	83	0101 0011	~	117	0111 0101
T	84	0101 0100	~	118	0111 0110
U	85	0101 0101	~	119	0111 0111
V	86	0101 0110	~	120	0111 1000
W	87	0101 0111	~	121	0111 1001
X	88	0101 1000	~	122	0111 1010
Y	89	0101 1001	~	123	0111 1011
Z	90	0101 1010	~	124	0111 1100
[91	0101 1011	~	125	0111 1101
\	92	0101 1100	~	126	0111 1110
]	93	0101 1101			

6

Coder des images

Un scanner ou un appareil photo :

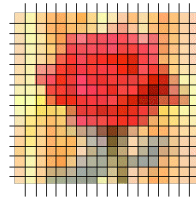
- Décompose une image en points : les pixels. (Premier critère de qualité : nombre de px)
- Associe une couleur à chaque pixel.
- Décompose cette couleur en 3 composantes : Rouge, Vert, Bleu.
- Note l'intensité de chaque composante par un nombre entre 0 et 255.

Exemple : (192, 64, 255) est un pixel violet

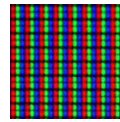
Avec ce codage 1 pixel se note sur 3 octets.

7

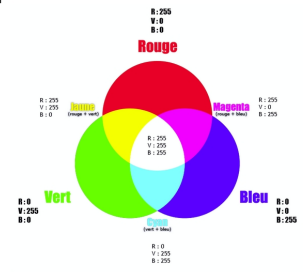
Le codage RVB



Poids brut d'une image = 3*nombre de pixels



Votre moniteur (très agrandi)



8

Différents formats

Le rangement des données numérisées dans un fichier nécessite le choix d'une méthode de rangement :

c'est le **FORMAT** du fichier

- Pour l'utilisateur, il est représenté par une **extension du nom de fichier** (2, 3 ou 4 caractères terminant le nom du fichier et précédés d'un point)

Ex : Format Word \Rightarrow extension = **.docx**

9

Les formats sont gérés par les logiciels :

- Le format est déterminé **par le logiciel** au moment **de l'enregistrement** d'un document.
- Chaque logiciel connaît un nombre limité de formats.
- Pour **CHANGER** le format d'un fichier il **FAUT** donc passer par un logiciel **connaissant le format d'origine et le format final**.
- Le logiciel système associe (par défaut) un logiciel à chaque format.

10

Principaux formats pour les textes

Tout dépend de ce que l'on veut enregistrer dans le fichier :

- Le **contenu seulement** (texte brut)
 - Format dit TEXTE : txt
- **+les enrichissements** (police, corps, couleurs, ...)
 - Format TEXTE ENRICHIS: rtf
- **+fonctions avancées** (pagination, tables, index, ...)
 - Doc, docx, odt

Le pdf est un format particulier : c'est un format pour la distribution de documents terminés, pas un format de travail.

11

Principaux formats : Images

Tout dépend de la qualité finale souhaitée et du poids accepté :

- **Gif** : peu de couleurs (256), peu de poids (quelques Ko)
- **Jpg** : beaucoup de couleurs (millions), peu de poids (le plus souvent moins de 1Mo)
 - On choisit le niveau de qualité en fonction du but final (photo d'identité ou poster)
 - Compression jpg : supprimer les px superflus.
- **Tif** : format de travail (beaucoup de couleurs et de pixels (poids : dizaines de Mo)
 - Pas de suppression de px par rapport à la prise de vue

12

Principaux formats : audio

- **Codage** : Numérisation par échantillonnage. (Premier critère de qualité)
- CD audio = qualité moyenne (échantillonnage : 16bits, 44KHz) : **10Mo/min en stéréo**.
- **Le son** peut ensuite être compressé (perte de qualité par rapport à l'échantillonnage).

Formats courants (son compressé, poids : autour de 2Mo/min) :

- AAC (Apple : itunes)
- MP3 (...)
- WMA (Windows : windows media)
- OGG (format libre)

13

Principaux formats : vidéo

Codage : La taille d'une vidéo numérisée image par image est énorme → compression (avec perte) obligatoire pour la distribution.

Méthode de compression = CODEC :

- Webm, H264, Ogg sont des codecs utilisés aujourd'hui.

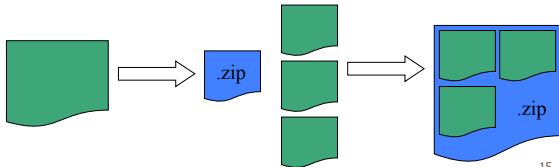
La vidéo compressée est enregistrée dans un fichier; de nombreux formats existent :

- .mp4
- .mkv
- .wmv (Microsoft : windows media player)
- .mov (Apple : quicktime)

14

Principaux formats : archivage

- **Compression d'archivage** = réduire la taille en appliquant une méthode qui enregistre l'information sous une forme plus compacte mais **sans perte**
- **Archivage** = regrouper plusieurs fichiers **dans un seul**/ regrouper des répertoires, des sous-répertoires et tout leur contenu **dans un seul fichier**



15