

FORMATS DE FICHIERS

Stage – Semaine 5

FORMATS DE FICHIERS

- Déjà vu : il existe plusieurs formats d'images ?
 - à cause de choix de codage différents
 - Codage matriciel versus codage vectoriel
 - Codage de couleurs (niveaux de gris, modèle RVB, ...)
 - Utilisation ou non d'algorithme de compression (potentiellement destructrice)
- La multitude de formats existe pour d'autres types de documents numériques
 - La raison est toujours des codages différents...
 - ... en suite de 0 et de 1, le bit étant l'information de base pour un ordinateur

CODAGE ?

Codage d'une information :

Règle d'écriture à l'aide d'un nombre fini de symboles permettant de désigner de manière unique l'information

(le codage est contextualisé (codage de nombre, de couleurs, de textes, ...))

Exemple : codage d'entiers

123 ↔ Cent vingt trois ↔ CXXIII

Trois écritures de la même valeur

à l'aide d'une liste de symboles indécomposables :

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

Trois niveaux de formats dépendant du type de contenu

- **contenu seulement** (texte brut)
 - Format dit TEXTE : txt
 - Texte formaté par l'utilisateur : html, xml, ...
 - Programmes informatiques : php, c, ...
- **+ les enrichissements** (police, corps, couleurs, ...)
 - Format TEXTE ENRICHIS: rtf
- **+ fonctionnalités avancées** (pagination, tables, index, ...)
 - Format de travail de bureautique : doc, docx, odt
 - Format d'échange et de distribution : pdf

LE FORMAT TEXTE SEULEMENT

- La seule information qu'il contient = une suite de caractères
- Conseillé d'utiliser comme logiciel, un « éditeur de texte » (TextWrangler, BBedit, emacs, NotePad++, ...) et non un « logiciel de traitement de texte » (TextEdit, Open Office Writer, Libre Office Writer, Microsoft Word, ...)
- Chaque caractère est codé par (i.e. associé à) une suite de 0 et de 1

A

0	1	0	0	0	0	0	1
---	---	---	---	---	---	---	---

Exemple :

codage de A et B en
code ASCII ou UTF-8

B

0	1	0	0	0	0	1	0
---	---	---	---	---	---	---	---

- ... mais plusieurs tables de codage possibles
 - ASCII
 - Unicode (utf-8, ...)
 - ...
(voir diapositive suivante)

DE L'ASCII À L'UTF-8

- **ASCII d'origine** (aperçu à droite)
 - 7 bits par caractère : caractères latins sans accent.
 - 128 caractères
- ASCII étendu
 - **nombreuses extensions** non compatibles entre elles (une par langue)
 - 8bits par caractère
 - 256 caractères
- « utf-8 » :
 - **une table universelle** (recommandée W3C)
 - Codage variant de 8 à 32 bits pour un caractère ou idéogramme

盲人摸象 - 各執一端

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	a	97	0110 0001
\$	36	0010 0100	b	98	0110 0010
A	65	0100 0001	c	99	0110 0011
B	66	0100 0010	d	100	0110 0100
C	67	0100 0011	e	101	0110 0101
D	68	0100 0100	f	102	0110 0110
E	69	0100 0101	g	103	0110 0111
F	70	0100 0110	h	104	0110 1000
G	71	0100 0111	i	105	0110 1001
H	72	0100 1000	j	106	0110 1010
I	73	0100 1001	k	107	0110 1011
J	74	0100 1010	l	108	0110 1100
K	75	0100 1011	m	109	0110 1101
L	76	0100 1100	n	110	0110 1110
M	77	0100 1101	o	111	0110 1111
N	78	0100 1110	p	112	0111 0000
O	79	0100 1111	q	113	0111 0001
P	80	0101 0000	r	114	0111 0010
Q	81	0101 0001	s	115	0111 0011
R	82	0101 0010	t	116	0111 0100
S	83	0101 0011	u	117	0111 0101
T	84	0101 0100	v	118	0111 0110
U	85	0101 0101	w	119	0111 0111
V	86	0101 0110	x	120	0111 1000
W	87	0101 0111	y	121	0111 1001
X	88	0101 1000	z	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			

RECONNAISSANCE DE FORMATS

- **Extension de fichiers** : suffixe de 2 à 4 lettres indiquant le format utilisé pour coder l'information contenue dans le fichier
 - Exemples : txt, odt, docx, pdf, ..., jpg (pour format jpeg), ...
- Le logiciel système associe un logiciel à chaque format (le logiciel associé par défaut peut-être modifié).
- Si le suffixe associé à un fichier est erroné (ne correspond pas au format réellement associé), le logiciel par défaut risque de ne pas savoir ouvrir le document → fichier inutilisable
- Conséquence : conseil =
laisser les logiciels ajouter eux-mêmes les extensions
(Ne saisissez que les noms hors suffixes)
car ils connaissent l'extension adéquate pour le format qu'ils utilisent

TRANSFORMATION DE FORMATS

- Changer (ou saisir) un suffixe ne change pas le format réel du contenu d'un fichier.
- Par contre, les logiciels connaissent généralement :
 - Plusieurs formats
 - Les traitements pour passer de l'un à l'autre (via l'enregistrement ou l'export du fichier... en précisant le « type de fichier » ou le « format de fichier »)
- Ces fonctionnalités d'enregistrement ou d'export sont à utiliser pour changer le format d'un fichier.