

Chapitre 3 : Structure et codification de l'information

I- LA STRUCTURE DE L'INFORMATION

L'information peut être représentée par différentes formes :

- **Le caractère** : une lettre, un chiffre, un symbole
- **Le mot**: assemblage de plusieurs caractères qui prennent une signification.
- **La rubrique**: assemblage de plusieurs mots qui prennent un sens. la rubrique est le plus petit ensemble d'information qui puisse être utilisé de façon autonome.
- **L'article**: ensemble de rubriques représentant les propriétés d'un objet ; l'article permet son identification.

II- CODIFICATION

La codification des données permet :

- √ D'identifier avec précision chaque donnée
- √ De condenser l'information
- √ D'accélérer le traitement

1- Codification fonctionnelle :

a- Définition

La codification est une opération intellectuelle qui correspond à une forme symbolique de la structure de l'information.

Ainsi, un code est une représentation conventionnelle abrégée d'une information :

- Il identifie l'objet sans ambiguïté
- Il permet de réaliser :
 - des gains de place sur les supports informatiques ;
 - des gains de temps lors de la saisie de l'information.
- Il permet de contrôler l'information.

b- Les différents types de codes

- **Un code élémentaire est élaboré à partir d'une seule propriété**

Code	Mode d'élaboration	Appréciation
Séquentiel	Numéro attribués en suivant : Ex : matricule	+ simple, non ambigu - Non significatif
Significatif descriptif	Lettres ou chiffres représentant une caractéristique de l'objet : Ex : taille 40	+ significatif - Insuffisant, en général, pour définir un objet (souvent associé à un code séquentiel) + Facile à mémoriser et à contrôler

➤ Un code complexe traduit plusieurs propriétés

Juxtaposé	Code formé de tranches indépendantes Ex : code à barres 13 caractères : 1 pour le pays, 5 pour le fabricant ; 6 pour le produit, 1 pour le contrôle	+ Non ambigu Possibilité de contrôles, de statistiques. - Long, lourd
Hiérarchique	Code dans lequel il existe une subordination entre les tranches qui le composent. Ex : code du plan comptable	+ Non ambigu - Long, lourd

➤ Un code important est souvent assorti d'une clé de contrôle

Les codes contrôlables	Méthode arithmétique « MODULO 10 »		
	Pour obtenir la clé de contrôle, chaque chiffre est multiplié par 2 ou par 1 en partant de la gauche selon son rang pair ou impair. La somme des produits est retranchée de la dizaine immédiatement supérieure.		
	Code	Calcul de la clé	Code de contrôle
	3261	$(3 \times 2) + (2 \times 1) + (6 \times 2) + (1 \times 1) = 21$ $30 - 21 = 9$	32619 ↑ Clé de contrôle
Méthode de la lettre de contrôle « MODULO 23 »			
Pour obtenir la clé de contrôle. Le code est divisé par 23. Le reste correspond à une lettre de prise dans une table.			
Code	Calcul de la clé	Code de contrôle	
3261	$\begin{array}{r} 3261 \\ 23 \overline{) 3261} \\ \underline{96} \\ 041 \\ \underline{18} \\ 0 \end{array}$ <p style="text-align: center;">0 1 2 3 ... 18 A B C D ... V</p>	3261 V → Clé de contrôle	
Remarque : les lettres I, O et S ne sont pas prises en considération pour ne pas être confondues avec les chiffres 1, 0 et 5			

2- Codification technologique :

La codification technologique appelée codage, consiste à traduire les informations constituées de chiffres de lettres et de signes en un langage directement assimilable par la machine. En effet, le nombre important et la diversité des caractères du langage humain (lettres, chiffres, signes...) justifient leur codage.

2-1 Le système binaire :

Il est construit autour d'un système numérique à base dans lequel les nombres sont exprimés sous formes de combinaisons de 0 et 1. La plus petite unité de mémoire de l'ordinateur est le BIT (Binary Item) qui peut prendre les valeurs 0 ou 1 .

Systeme decimal	0	1	2	3	4	5	6	7	8	9
Systeme binaire	0	1	10	11	100	101	110	111	1000	1001

Les fichiers binaires sont formés d'une suite d'octets (un octet = 8bits) .

2-2- Les codes binaires

Les codes le plus souvent utilisés par les machines pour représenter l'information sont

- Le code ASCII (American Standard Information Interchange) :
 - Mots de 8 bits constitués de 7 éléments binaires (le huitième était réservé au contrôle).
 - Permet de représenter 128 caractères distincts.
 - Reconnu au niveau international.

- Le code EBCDIC (Extended Binary Coded Decimal Interchange Code) :
 - Mots de 8 bits plus 1 de contrôle.
 - Utilisation d'un demi-octet pour les chiffres décimaux (soit 4 bits).
 - Permet de représenter 256 caractères distincts.
 - Utilisés par la plupart des ordinateurs.