

LA CODIFICATION

I/ OBJECTIFS

Le problème de la codification est un problème de gestion des informations .Il devient très important avec le développement des traitements automatisés. Aussi, l'une des préoccupations des entreprises est la **mise en œuvre de nouveaux codes et le contrôle des codes**.

La codification est une opération particulière d'optimisation : elle substitue à une information sous forme « naturelle » un symbole mieux adapté aux objectifs de l'utilisateur. D'où, à un symbole ne doit correspondre qu'un seul objet représenté tel qu'il n'y aura ni synonymes ni polysèmes.

La codification porte sur le nom de l'information à codifier et sur sa valeur. Par exemple :

Nom de l'information : Situation familiale : Code: SIT-FAM

Valeur de l'information: Marié(e) : M ; Célibataire : C ; Veuf(ve) : V ; Divorcée : D

La codification permet:

- De gagner du temps lors des opérations de lecture et d'écriture (qu'elles soient automatisées ou non) et de la place sur les supports (les imprimés ou les supports automatisés).

Ex.: il vaut mieux désigner un livre par son code au lieu de son titre.

- De détecter certains types d'erreurs. Pour une classe d'objets, le format du code est identique (même nombre de caractères).

Ex.: Le matricule d'un véhicule algérien se compose d'un nombre entier suivi du type du véhicule (un chiffre compris entre 1 et 8) et de l'année de mise en circulation du véhicule (2 chiffres), suivi du code de la wilaya d'immatriculation du véhicule (2 chiffres). Ces trois nombres sont séparés par un espace (blanc). Ainsi le matricule 5208 11 19 est faux alors que le matricule 5208 114 19 est juste.

- De détecter certaines erreurs de transcription (recopie d'un code d'un support sur un autre). Par exemple quand on recopie un numéro de compte bancaire on risque d'oublier des chiffres ou de les permuter.

- De reconnaître certaines propriétés de l'objet et de définir certains traitements communs à des objets ayant la même valeur de code, lorsque le code est bien conçu.

Ex.: Si dans le fichier « CLIENTS » le 6 ème caractère du code client correspond au mode de paiement du client avec les valeurs:

1 = règlement par chèque;

2 = règlement en espèces;

il sera facile de sélectionner les clients réglant par chèque (Ex: le client 412271).

II/ DIFFERENTS TYPES DE CODES

1. Code séquentiel

Un nombre pris dans la suite des entiers consécutifs est attribué à chaque objet quand il se présente.

Ex.: Le dernier client a le code 1437, le suivant aura le code 1438.

- Avantages :

- strictement non ambigu;
- simple à mettre en œuvre.

- Inconvénients:

- non significatif;
- ne permet pas d'insertion d'une valeur entre deux anciennes.

2. Code séquentiel par tranches

On réserve une suite de numéros consécutifs pour un groupe de données possédant une propriété commune. On laisse des numéros libres à l'intérieur de la tranche pour maintenir le classement après insertion de données nouvelles. On codifie séquentiellement à l'intérieur de la tranche.

Ex.: Les clients détaillants seront codés entre 1 et 499 et les grossistes entre 500 et 600.

- Avantages :
 - non ambigu si aucun objet n'appartient à plusieurs tranches à la fois;
 - permet des insertions.
- Inconvénients:
 - difficultés de définir la dimension des tranches;
 - signification du code très limitée.

3. Codes significatifs

3.1. Code mnémotechnique

Les symboles utilisés facilitent la mémorisation de la signification parce qu'ils sont évocateurs des objets désignés.

Ex.: FACT ou FCTR pour facture, TOT pour TOTAL, DAT-INSC pour DATE D'INSCRIPTION.

Il est très prudent de tester les réactions de utilisateurs avant la mise en service du code parce qu'un code peut être interprété différemment.

Ex.: DP peut être traduit par Direction de la Production ou Direction du Personnel.

3.2. Code descriptif ou analytique

Les symboles retenus intègrent des caractéristiques permanentes l'objet codé.

Ex.: Dans une nomenclature industrielle, VR 452 désigne des vis à tête ronde, de longueur de 45 mm, au pas de 2 mm.

- Avantages :
 - pratiques pour les utilisations manuelles car simples à comprendre;
 - extensibles, car à de nouveaux objets correspondent toujours les valeurs nouvelles des propriétés retenues.
- Inconvénients :
 - risques d'ambiguïté si le nombre et la nature des propriétés retenues ne permettent pas de distinguer, dans tous les cas, deux objets différents.

4. Codes juxtaposés (concaténés) ou articulés

Les codes sont découpés en plusieurs zones concaténées. Chaque zone correspond à une propriété particulière de l'objet codé.

Ex.1: Dans la codification d'un salarié d'une entreprise, on a retenu les propriétés catégorie et atelier complétées par un numéro séquentiel : 19 06 245. 19 désigne la catégorie et 06 l'atelier.

Ex.2: Numéro d'identification des individus par l'INSEE français. Ce numéro à 13 chiffres (plus 2 de contrôle) juxtapose plusieurs propriétés dont l'ensemble permet une identification non ambiguë de chaque français. Il est codé de la manière suivante :

Sexe + deux derniers chiffres de l'année de naissance + mois de naissance + N° de la commune dans le département + N° d'ordre du registre de naissance de la commune.

Par exemple : 2 48 03 31 107 021 est une femme (1 sexe masculin, 2 sexe féminin) née en 1948 au mois de mars dans le département 31 dans la commune 107 du département 31 et a le numéro 021 sur le registre de naissance de la commune.

- Avantages:
 - identification non ambiguë
 - éventuellement, possibilités de contrôle sur certaines propriétés.

- Inconvénients:
 - codification « lourde » pouvant nécessiter un nombre important de caractères;
 - nécessiter de choisir des propriétés stables sinon le code n'est plus totalement significatif.

5. Codes hiérarchisés ou à niveaux

Ils sont utilisés dans le cas où il existe des relations d'inclusion entre les différents ensembles auxquels appartient l'objet. Le code comporte plusieurs zones concaténées. Généralement, les zones situées à droite désignent des ensembles inclus dans les ensembles désignés par les zones de gauche.

Ex.1: Le code postal se compose ainsi: 2 chiffres indiquent la wilaya, le 3ème chiffre l'APC dans la wilaya et les deux derniers chiffres le N° d'arrondissement dans l'APC si des arrondissements existent dans cette APC.

Ex.: le 4^{ème} arrondissement de l'APC d'El-Eulma a pour code postal 19604.

Ex.2: Le code des adresses aériennes est construit sur un découpage en continent (1 chiffre), zone (1 chiffre), pays (2 chiffres), ville (2 chiffres).

Ex.3: les comptes du plan comptable expriment la classe comptable, le chapitre, etc.

Avantages :

- très bonne aptitude au regroupement et à la partition;
- relativement simple à concevoir.
- Inconvénients :
 - lourdeur du code nécessaire pour une codification non ambiguë.

III/ LES CONTROLES

L'objectif des opérations de contrôle est de vérifier l'exactitude et la conformité de l'information. Exemple. Numéro étudiant dont la spécialité n'existe pas.

Les contrôles peuvent être directs c'est à dire effectués sur l'information elle-même sans tenir compte des autres informations du système, ou indirects ce qui revient à vérifier la conformité de l'information par rapport à d'autres informations du système.

1. Codes contrôlables

Le principe consiste à associer à une valeur du code une *clé* de contrôle (chiffres ou lettres) dont la valeur est calculée à partir de la valeur du code et vérifiée à chaque utilisation pour éviter les erreurs d'identification qui peuvent avoir des conséquences graves.

Ex.: Effectuer une opération bancaire sur le compte 552635 au lieu du compte 552653 à cause d'une erreur à l'écriture ou à la recopie. Convoquer le malade N° 123 au lieu de 132, ...

Les méthodes, pour associer une clé à un code sont très nombreuses. On peut en créer selon les besoins. Nous citerons quelques exemples classiques.

1.1. Méthode arithmétique « modulo 10 »: Chaque chiffre du code est multiplié par 2 ou 1 selon son rang, pair ou impair; on additionne les produits partiels et on retranche de la dizaine immédiatement supérieure pour obtenir le chiffre le chiffre clé correspondant.

Ex.: Code d'origine 02456 donnera $(0*1)+(2*2)+(4*1)+(5*2)+(6*1) = 24$

La dizaine supérieure à 24 est 30. Le chiffre-clé est $30 - 24 = 6$.

Le code contrôlé sera le code d'origine auquel on concatènera le chiffre clé soit : 024566.

Si par exemple on saisit 024563, le code sera rejeté car il y a eu erreur sur le chiffre clé.

1.2. Méthode géométrique « modulo 11 »: On multiplie chaque chiffre du code par les puissances successives de 2 et on additionne les produits partiels. La somme est divisée par 11. Le reste de la division est utilisé comme chiffre nombre-clé.

Ex.: Code d'origine : 02456 donnera: $(0*32)+(2*16)+(4*8)+(5*2)+(6*1) = 80$

Le reste de $80 / 11$ est égal à 3. Le code contrôlé sera 024563.

1.3. Méthode de la lettre de contrôle « modulo 23 »: Le code d'origine est divisé par 23. Au reste de la division on fait correspondre une lettre-clé prise dans une table établie auparavant. On élimine de la table les lettres I, O et S pour éviter les confusions avec respectivement les chiffres 1, 0 et 5.

Ex.: Code d'origine : 02456. Le reste de la division $02456 / 23$ est 18 qui correspond à la lettre A dans la table.

Soit la table des lettres qu'on a réalisé au préalable:

0	1	2	3	4	5	6	7	8	9	10	11
F	G	H	J	K	L	M	N	P	Q	R	T
12	13	14	15	16	17	18	19	20	21	22	
U	V	W	X	Y	Z	A	B	C	D	E	

Le code contrôlé sera, en utilisant, cette table 02456A.

On peut par exemple, avant de diviser par 23, faire la multiplication des chiffres du code d'origine par les puissances successives de 2.

Remarque 1: Ces codes auto-correcteurs d'erreurs sont utilisés avec des ordinateurs et certaines erreurs ne sont pas détectées par ce procédé.

Remarque 2 : Lorsqu'un code est alphanumérique, avant de calculer la clé de contrôle, on remplace chaque lettre du code par son numéro (sa position) dans l'alphabet pour obtenir un code numérique.

Ex.: A = 1, B = 2, ... Le code 12X5 aura pour équivalent 12245 (X a été remplacé par 24). On applique ensuite une des méthodes de calcul de clé.

2. Contrôles directs

Les principaux contrôles directs sont :

- Contrôle de présence. Exemple : On doit saisir le nom de l'étudiant.
- Contrôle de type. Exemple : le nom de l'étudiant ne doit pas contenir de chiffres
- Contrôle de cadrage (position de l'information dans la zone de saisie). Exemple : le nom de l'étudiant doit commencer à la 1^{ère} case à gauche.

3. Contrôles indirects

Parmi les contrôles indirects, on peut citer :

- Contrôle de cohérence interne. Exemple : Numéro d'inscription de l'étudiant : D03/13/0503 est correct mais le numéro D30/13/0503 est faux car le domaine D30 n'existe pas
- Contrôle de cohérence externe (Domaine d'étude : D03 (Math et informatique) et Nom de la licence : Droit);
- Contrôle de vraisemblance (DAT-INSC = 32/09/13 est fausse car $\text{jour} \leq 31$ si mois différent de 02).

IV/ PRATIQUE DE LA CODIFICATION

La codification est une opération importante: un code mal adapté est une source d'erreurs et de retards dans les traitements. Les qualités à attendre d'un code sont :

- 1/ Concision: plus un code est court moins il y a d'erreurs de recopie et plus on gagne en temps de lecture et d'écriture.
- 2/ Facilité de contrôle: le code conçu doit être facilement contrôlable manuellement ou automatiquement.
- 3/ Facilité d'exploitation: si le traitement est manuel un code alphabétique est plus agréable. Dans le cas d'une machine à calculer le code numérique est le seul utilisable.
- 4/ Facilité de décodification: l'interprétation du code doit être facilitée au maximum.
- 5/ Unicité et sélectivité: Le code doit permettre d'isoler l'objet recherché sans ambiguïté dans l'ensemble auquel il appartient.
- 6/ Souplesse et pérennité: un code peut être utilisé très longtemps (plusieurs dizaines d'années parfois) c'est la pérennité. La souplesse et une qualité qui lui permet de s'adapter aux modifications de la population représentée et ainsi de conserver sa valeur descriptive au cours du temps. Le code doit permettre des extensions lorsque le nombre d'objets s'accroît c'est à dire qu'il doit être possible de coder tous les nouveaux objets.

Le code doit aussi permettre des insertions car dans de nombreux cas de nouveaux objets doivent être insérés entre les codes déjà attribués.

Ex.1: Si on a actuellement 50 employés on devra avant de choisir un code sur 2 positions voir l'évolution de l'entreprise ainsi que le taux de renouvellement du personnel.

Ex.2: Si on codifie nos clients par tranche « détaillant » et « grossiste » par exemple de 01 à 20 les clients grossistes et de 21 à 99 les détaillants, les tranches doivent être suffisantes pour coder les futurs clients.

Ex3.: La codification des wilayas en Algérie selon leur ordre alphabétique n'a pas permis d'insertion de nouveaux codes entre ceux déjà attribués. Ainsi lorsqu'il y a eu un nouveau découpage, la wilaya de Bordj Bou Arreridj a eu le code 34 alors que la lettre B de Bordj aurait dû être classée après les wilayas commençant par la lettre A.

IV/ CHOIX DE LA SOLUTION

Il est utile de tenir compte des recommandations suivantes:

- Avant de créer un nouveau code, il faut d'abord vérifier si des codes existant au sein de l'organisation elle-même ou chez des organismes spécialisés ne peuvent être utilisés (institut des statistiques ou de normalisation industrielle).
- Consulter tous les futurs utilisateurs avant de choisir la nature du code.
- Vérifier la stabilité des propriétés sur lesquelles sera basé le code. Par exemple, le sexe d'une personne est une propriété stable qui peut figurer dans un code.
- Vérifier si la capacité du code permet de faire face aux extensions et aux insertions prévisibles.
- Prévoir une procédure de mise en œuvre des nouveaux codes (information, formation, suppression des codes périmés, date de mise en application du nouveau code, ...).
- Prévoir des contrôles sur les documents de bases et des contrôles automatiques.