

Leçon 3 - Représentation de l'information

JOËL ADÉPO, UVCI

Octobre 2017

Table des matières

Objectifs	3
I - Codage de l'information	4
A. Types d'information.....	4
B. Codage de l'information.....	4
C. Exercice.....	7
II - Numérisation de l'information analogique	9
A. Principe de numérisation.....	9
B. Exercice.....	10
Solution des exercices	12

Objectifs

A la fin de cette leçon, vous serez capable de :

- Identifier les différents types d'informations
- Décrire les principales techniques de codage de l'information
- Décrire les méthodes de numérisation de l'information.

Codage de l'information



A. Types d'information

Les informations transmises peuvent être réparties en deux grandes catégories selon :

- l'élément qu'elles représentent
- et les traitements subis dans les systèmes informatiques.

On distingue :

- les données discrètes,
- les données analogiques ou continues.

1. Données discrètes

Dans les données discrètes, l'information correspond à l'assemblage d'une suite d'éléments indépendants les uns des autres (suite discontinue de valeurs) et dénombrables (ensemble fini).

Par exemple, un texte est une association de mots eux-mêmes composés de lettres (symboles élémentaires).

2. Données continues

Les données continues ou analogiques résultent de la variation continue d'un phénomène physique : température, voix, image...

Un signal analogique est un signal qui varie de manière analogue au phénomène physique.

Un signal analogique peut prendre une infinité de valeurs dans un intervalle déterminé (bornes).

Le traitement des informations par des équipements informatiques nécessite de remplacer à chaque élément d'information une valeur binaire.

On parle alors de :

- **codage de l'information** (codage à la source) pour les informations discrètes
- et **numérisation** de l'information pour les informations analogiques.

B. Codage de l'information

Définition : Coder une information

Coder l'information consiste à faire correspondre (bijection) à chaque symbole d'un alphabet (élément à coder) une représentation binaire (mot code). L'ensemble des mots codes constitue le code.

Les informations à coder peuvent être des commandes d'une machine-outil ou des caractères alphanumériques...

Les informations que nous codons dans la suite sont des caractères alphanumériques.

1. Différents types de code

Le codage des différents états d'un système peut s'envisager selon deux approches. La première, la plus simple, considère que chacun des états du système est équiprobable. La seconde prend en compte la fréquence d'apparition d'un état. Cette approche conduit à définir deux types de code :

- les codes de longueur fixe ,
- et les codes de longueur variable.

2. Codes de longueur fixe

Chaque état du système est codé par un certain nombre de bits, appelé longueur du code, longueur du mot code ou encore code à n moments.

Avec 1 bit on peut coder 2 états (0,1)

- Avec 2 bits on peut coder 4 états (00, 01, 10, 11)
- Avec 3 bits on peut coder 8 états (000, 001, 010, 011, 100, 101, 110, 111).

D'une manière générale :

Avec n bits on code 2^n états (^ est le symbole d'exposant).

Définition : 2.1. Puissance lexicographique

Le nombre d'états pouvant être codés par un code de n bits s'appelle **puissance lexicographique** du code .

La puissance lexicographique se note P:

$$P=2^n.$$

Le nombre de bits pour coder P symboles est donnée par l'expression suivante :

$$n = \log_2(P)$$

2.2. Quantité d'information

Ce nombre de bits (n) représente la quantité d'information (Q) apportée par la connaissance d'un état du système.

Lorsque dans un système, **tous les états sont équiprobables**, la quantité d'information apportée par la connaissance d'un état est la même quel que soit l'état connu.

Si l'information est représentée par deux valeurs équiprobables (0 ou 1, pile ou face...), la quantité d'information, exprimée en bit, est $Q=1$ bit (le logarithme à base 2 de 2 est égal à 1).

Lorsque tous **les états ne sont pas équiprobables**, la quantité d'information est d'autant plus grande que la probabilité de réalisation de l'état est faible. Si p est la probabilité de réalisation de l'état E, la quantité d'information apportée par la connaissance de P est donnée par l'expression suivante :

$$Q = \log_2\left(\frac{1}{p}\right)$$

3. Codes de longueur variable

Lorsque les états du système ne sont pas équiprobables, la quantité d'information apportée par la connaissance d'un état est d'autant plus grande que cet état a une faible probabilité de se réaliser.

3.1. Quantité moyenne d'information

La quantité moyenne d'information apportée par la connaissance d'un état, appelée entropie, est donnée par la relation suivante :

$$H = \sum_{i=1}^{i=n} p_i \log_2 \frac{1}{p_i}$$

p_i représente la probabilité d'apparition du symbole de rang i .

L'entropie représente la longueur optimale du codage des symboles du système.

Exemple

Déterminons la longueur optimale du code pour le système décrit par le tableau ci-dessous.

État	Probabilité
E	0,48
A	0,21
S	0.12
T	0.08
U	0.06
Y	0.05

Le code optimal utile déterminé avec la formule 1 est de 1,92 bit, alors que l'utilisation d'un code à longueur fixe nécessite 3 bits pour coder les 6 états de ce système ($2^2 < 6 < 2^3$).

3.2. Codage de Huffman

Le codage de Huffman prend en compte la fréquence d'apparition des états et se rapproche de la longueur optimale.

Construction du code de Huffman:

1. lecture complète du fichier et création de la table des symboles ;
2. classement des symboles par ordre des fréquences décroissantes (occurrence);
3. réductions successives en rassemblant en une nouvelle occurrence les deux occurrences de plus petite fréquence ;
4. l'occurrence obtenue est insérée dans la table et celle-ci est à nouveau triée par ordre décroissant;
5. les réductions se poursuivent jusqu'à ce qu'il n'y ait plus d'élément ;
6. construire l'arbre binaire en reliant chaque occurrence à la racine ;
7. le codage consiste à lire l'arbre du sommet aux feuilles en attribuant par exemple la valeur 0 aux branches basses et 1 aux branches hautes.

Exemple : Codage de Huffman

Codage de Huffman du tableau 1.

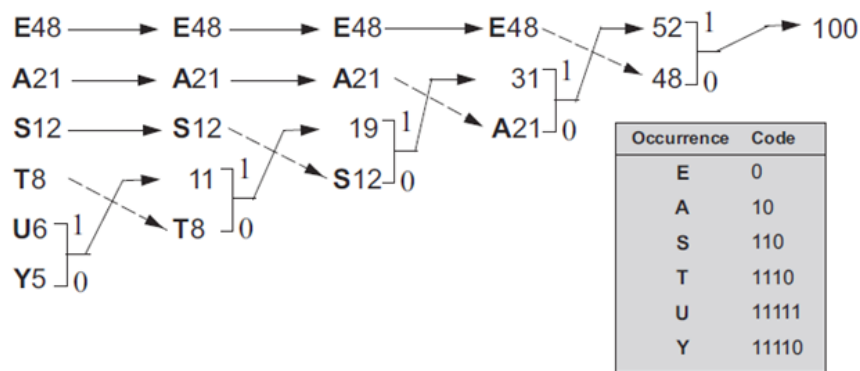


Figure 1 - Exemple de codage de Huffman

Les symboles U et Y ont été associés pour donner un autre symbole dans la deuxième colonne avec une fréquence qui est la somme des fréquences de U et Y. Ce même principe est répété pour obtenir les autres colonnes. U6 signifie que le symbole U a une fréquence de 0,06 ou 6%. On remarque dans chaque colonne que les symboles sont rangés dans l'ordre décroissant selon leur fréquences.

La longueur moyenne du codage ci-dessus est :

$$0,48 \times 1 + 0,21 \times 2 + 0,12 \times 3 + 0,08 \times 4 + 0,06 \times 5 + 0,05 \times 5 = 2,13.$$

Le codage de Huffman permet de réduire le nombre de bits utilisés pour coder l'information. Dépendant du contexte, il impose, avant la transmission, d'établir une convention ou de transmettre, avant les données, le contenu de la table construite par l'émetteur.

C. Exercice

[Solution n°1 p 13]

Question 1

1) Combien de bits sont nécessaires pour coder toutes les lettres de l'alphabet français?

- ☐ a. 4
- ☐ b. 5
- ☐ c. 6
- ☐ d. 26

Question 2

Avec le code ASCII à 7 moments, on peut représenter

- ☐ A) 64 caractères
- ☐ B) 128 caractères
- ☐ C) 256 caractères

Question 3

Soit le tableau suivant contenant les états d'un codage avec leur probabilité d'apparition. Donner le code de Huffman correspondant.

Etats	Probabilités
E1	0,30
E2	0,15
E3	0,06
E4	0,14
E5	0,20
E6	0,08
E7	0,07

Etats	Probabilités	codes
E1	0,30	11
E2	0,15	101
E3	0,06	0110
E4	0,14	001
E5	0,20	00
E6	0,08	0110
E7	0,07	1110

Tableau 1

Etats	Probabilités	codes
E1	0,30	11
E2	0,15	101
E3	0,06	0110
E4	0,14	100
E5	0,20	00
E6	0,08	0110
E7	0,07	0111

Tableau 2

Etats	Probabilités	codes
E1	0,30	000
E2	0,15	001
E3	0,06	010
E4	0,14	011
E5	0,20	100
E6	0,08	101
E7	0,07	110

Tableau 3

- ☐ a. Tableau 1
- ☐ b. Tableau 2
- ☐ c. Tableau 3

Numérisation de l'information analogique



A. Principe de numérisation

Numériser une grandeur analogique consiste à transformer la suite continue de valeurs en une suite discrète et finie. À cet effet, on prélève, à des instants significatifs, un échantillon du signal et on exprime son amplitude par rapport à une échelle finie (quantification).

Le récepteur, à partir des valeurs transmises, reconstitue le signal d'origine. Une restitution fidèle du signal nécessite que soient définis :

- l'intervalle d'échantillonnage qui doit être une constante du système (fréquence d'échantillonnage);
- l'amplitude de l'échelle de quantification, celle-ci doit être suffisante pour reproduire la dynamique du signal (différence d'amplitude entre la valeur la plus faible et la valeur la plus forte);
- que chaque valeur obtenue soit codée.

La figure ci-dessous représente les différentes étapes de la numérisation du signal.

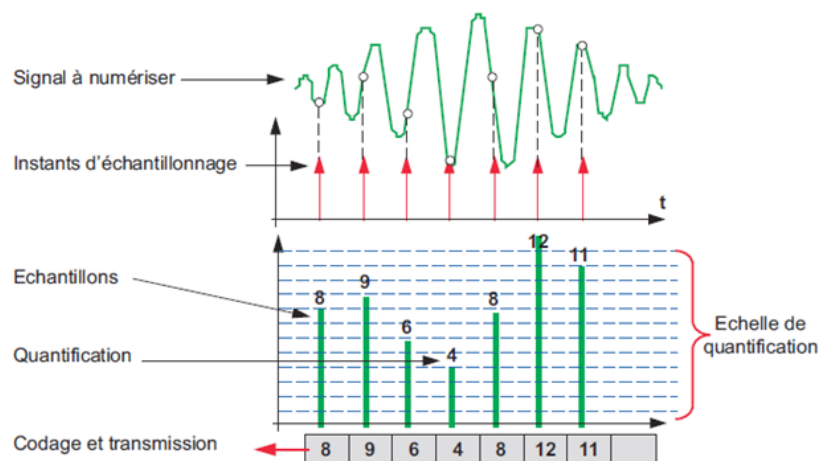


Figure 2 - Numérisation d'un signal analogique

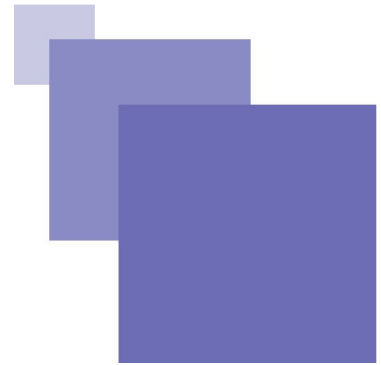
1. Échantillonnage

L'échantillonnage consiste à prélever à intervalle régulier (période d'échantillonnage), une fraction du signal (échantillon).

Plus la bande passante du signal est grande, plus il faut prendre d'échantillons par seconde.

- ☐ a. Une modulation
- ☐ b. une quantification puis un codage
- ☐ c. un codage
- ☐ d. une transmission

Solution des exercices



> Solution n°1 (exercice p. 7)

Question 1

- ☐ a. 4
- ☒ b. 5
- ☐ c. 6
- ☐ d. 26

Question 2

- ☐ A) 64 caractères
- ☒ B) 128 caractères
- ☐ C) 256 caractères

Question 3

- ☐ a. Tableau 1
- ☒ b. Tableau 2
- ☐ c. Tableau 3

> Solution n°2 (exercice p. 11)

Question 1

- ☐ a. 5 000 fois par seconde
- ☒ b. 10 000 fois par seconde
- ☒ c. 15 000 fois par seconde
- ☐ d. 8 000 fois par seconde

Question 2

- | | |
|----------------------------------|---------------------------------|
| <input type="radio"/> | a. toutes les secondes |
| <input checked="" type="radio"/> | b. toutes les 125 microsecondes |
| <input type="radio"/> | c. toutes les 125 millisecondes |

Question 3

- | | |
|----------------------------------|--------------------------------------|
| <input type="radio"/> | a. Une modulation |
| <input checked="" type="radio"/> | b. une quantification puis un codage |
| <input type="radio"/> | c. un codage |
| <input type="radio"/> | d. une transmission |