

Semestre : 6

Unité d'enseignement : UEF 3.2.2

Cours 3

**Codage et Théorie de
l'information**

Dr Mahmoud Hadeif

Codage de source

Partie 3

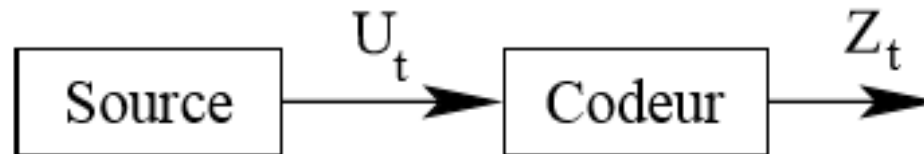
Codage de source

- **Codage efficace de l'information**
- Codage de Shannon-Fanno
- Algorithmes de Huffman

Codage de l'information

Définitions

- Coder pour compresser les données \rightarrow réduire (en moyenne) **la longueur des messages** \rightarrow tenter de supprimer un maximum de **redondance** des messages



- Une **source d'information** est un **générateur de messages** c'est-à-dire un générateur de séquences de symboles.
- Un **symbole** est simplement un élément d'un ensemble, appelé **alphabet**

Codage de l'information

Définitions

- Seuls les **alphabets finis** seront abordés → la source d'information est **discrète** → la taille de l'alphabet est appelée **l'arité** de la source.
- Seuls les messages de **longueur finie** seront considérés
- Les messages de la source entrent alors dans un **codeur** qui les transforme en une séquence de **mots de code**.
- Un **mot de code** est simplement une séquence de symboles pris dans **l'alphabet de codage**, un autre alphabet employé₄ par le codeur.

Codage de l'information

Définitions

- Chaque symbole a une **probabilité** $P(U_t = u_i)$ d'être émis par la source à l'instant t . On dit que la source est **sans mémoire** si la probabilité qu'un symbole u_i soit émis ne **dépend pas** des valeurs émises **précédemment**

$$\forall t \geq 1 \quad \forall u_i \in V_U \quad P(U_t = u_i | U_1 \dots U_{t-1}) = P(U_t = u_i)$$

- les **sources stationnaires** \rightarrow les sources pour lesquelles $P(U_t = u_i)$ **ne dépend pas de t** $\rightarrow P(U = u_i)$ sera dans la suite noté simplement p_i

Codage de l'information

Définitions

- le processus de codage $Z := f(U)$ est une application de l'alphabet de la source V_U à l'ensemble de mots de code V_Z
- Le code est **non-singulier** \rightarrow deux symboles différents de la source correspondent à deux mots de code différents

$$u_m \neq u_n \Rightarrow z_m \neq z_n$$

- **Exemple** \rightarrow code **Morse**. Il emploie essentiellement deux symboles : **un point (·)** et **un tiret (–)**. Par exemple, les lettres E, A et K sont respectivement codées «·», «· –» et «– · –».

Codage de l'information

Définitions

Le code **non-ambigu** \rightarrow si et seulement si chaque séquence (de longueur finie) de mots de code ne correspond qu'à un seul **message de la source**

Exemple (Code ambigu)

Considérons la source composée des trois symboles a, b et c. Le codage suivant de cette source

$a \rightarrow 1 \quad b \rightarrow 00 \quad c \rightarrow 11$ est ambigu.

une fois codés, il n'y a aucun moyen de distinguer le message «aaaa» du message «cc» deux sont codés «1111»

$a \rightarrow 1 \quad b \rightarrow 00 \quad c \rightarrow 10$ est non-ambigu

Codage de l'information

Définitions

- On dit qu'une séquence z de longueur n ($n \geq 1$) est un **préfixe** d'une autre séquence z' si et seulement si les **n premiers symboles** de z' forment exactement la séquence z

Exemple: **abba** est un préfixe de **abbabc**

- On dit que le code d'une source discrète est **sans préfixe** lorsqu'**aucun mot** de code n'est le préfixe d'un autre mot de code
- **Tout code sans préfixe est non-ambigu**

Codage de l'information

Définitions

Exemple

Considérons la source composée des trois symboles a, b et c.
Le code suivant de cette source :

$a \rightarrow 0$ $b \rightarrow 10$ $c \rightarrow 11$ est **sans préfixe**

D'autre part, le code suivant :

$a \rightarrow 1$ $b \rightarrow 00$ $c \rightarrow 10$ **n'est pas sans préfixe**

puisque 1 (le mot de code pour a) est un préfixe de 10 (le mot de code pour c).

Codage de l'information

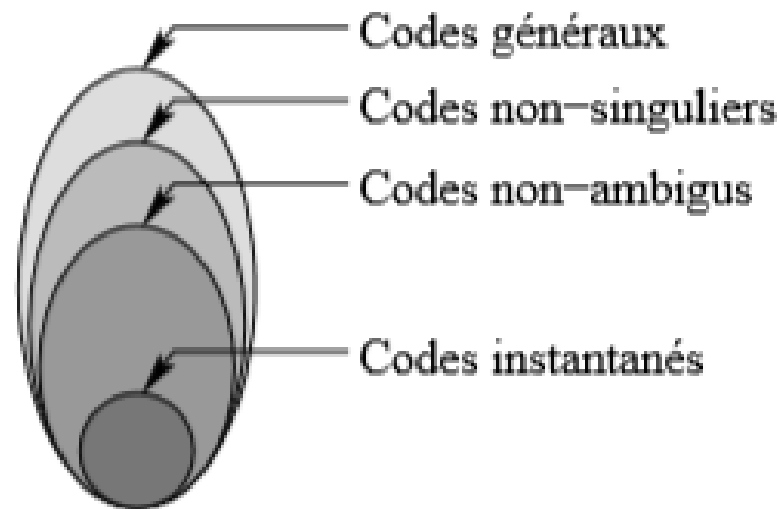
Définitions

- Un **code est instantané** → si et seulement si chaque mot de code dans toute chaîne de mots de code peut être **décodé dès que l'on a atteint sa fin**
- Un **code est instantané** si et seulement si il est **sans préfixe**
- **code est instantané** garantit qu'il n'est ni nécessaire de mémoriser les mots de code reçus ni d'attendre les suivants pour effectuer le décodage → permet d'économiser du **temps et de l'espace** dans le processus de décodage d'un message codé

Codage de l'information

Définitions

nous avons rencontré différents types de codes : **non-singulier**, **non-ambigu**, **instantané**. La façon dont ces différents types de codes sont reliés les uns aux autres est résumée



Codage de l'information

Exemple

Considérons une source d'information U , dont les symboles sont $u_1 = 1$, $u_2 = 2$, $u_3 = 3$, et $u_4 = 4$, avec la distribution de probabilité suivante :

$$P = \{ 0.5 \ 0.25 \ 0.125 \ 0.125 \}$$

Considérons donc le codage suivant de cette source, (où z_i est le mot de code pour u_i) :

$$z_1 = 0 \ , \ z_2 = 10 \ z_3 = 110 \ z_4 = 111$$

1. Le code est-il non-ambigu?
2. Codez le message 1423312.
3. Décodez la séquence 1001101010.

Codage de l'information

Exemple

Ces codes sont-ils sans préfixe? Non-ambigus? Instantanés?

a. $z_1=00$, $z_2=10$, $z_3=01$, $z_4=11$

b. $z_1=0$, $z_2=1$, $z_3=01$

c. $z_1=1$, $z_2=101$

Codage de l'information

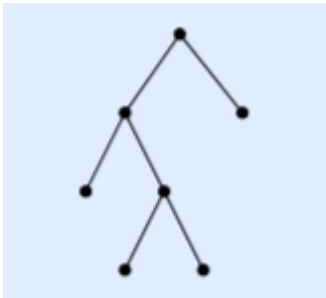
Définition et Propriété de l'Arbre n - aire

Définition: Un arbre n -aire ($n \geq 0$) est un arbre dans lequel chaque nœud intérieur a comme arité n , c'est-à-dire qu'il a exactement n fils. Un arbre n -aire complet est un arbre n -aire dans lequel toutes les feuilles ont la même profondeur.

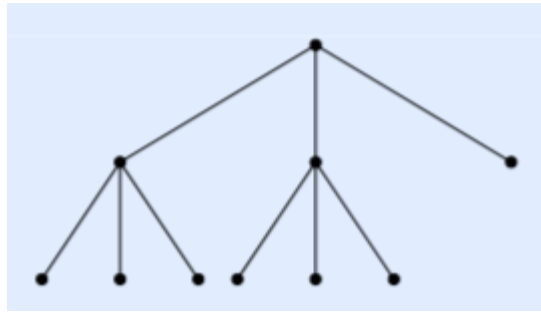
Propriété: Dans l'arbre n -aire complet de profondeur $d \geq 0$, chaque nœud à la profondeur δ ($0 \leq \delta \leq d$) couvre exactement $n^{d-\delta}$ feuilles.

Codage de l'information

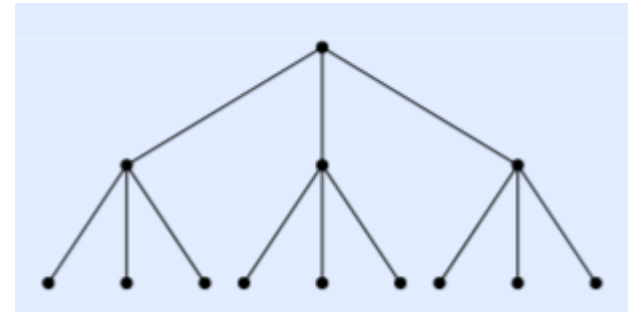
Exemple 4



Arbre binaire ($n = 2$)



Arbre binaire ($n = 3$)



Arbre ternaire complet

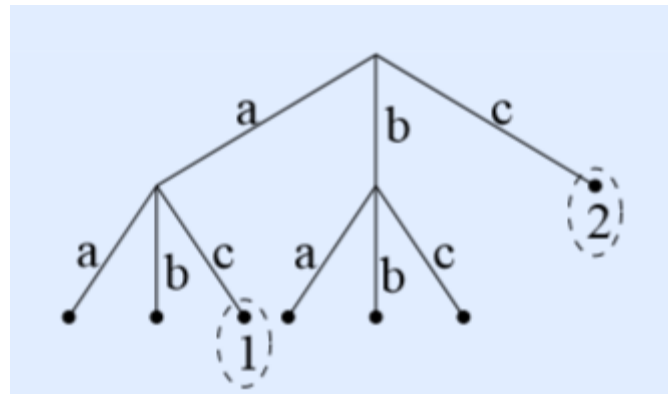
Codage de l'information

Définition de l'Arbre de Codage

Un arbre de codage est un arbre n -aire, dont les arcs sont étiquetés par des lettres d'un alphabet donné de taille n , de façon à ce que chaque lettre apparaisse tout au plus une fois à partir d'un nœud donné. Les mots de code définis par un tel arbre correspondent à des séquences d'étiquettes le long des chemins menant de la racine à une feuille.

Codage de l'information

Exemple 5



Un arbre de codage ternaire: le mot de code représenté par la feuille 1 est «**ac**» et la feuille 2 représente le mot de code «**C**».

Codage de l'information

Définition et Propriété du Code n -aire

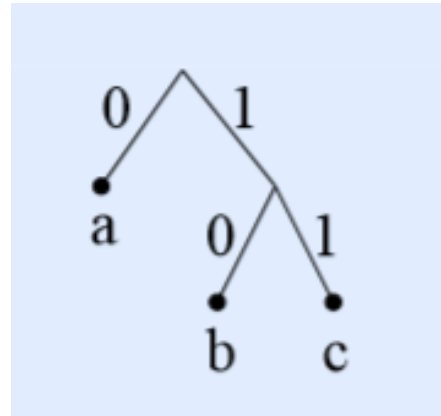
Définition: Un code avec un alphabet de taille n est appelé un code n -aire.

Propriété: Pour tout code instantané n -aire, il existe au moins un arbre de codage n -aire tel que chaque mot de code corresponde à la séquence d'étiquettes d'un chemin (unique) de la racine à une feuille. Inversement, chaque arbre de codage définit un code instantané. Les mots de code de ce code instantané sont définis comme les séquences d'étiquettes de chaque chemin de la racine à chaque feuille de l'arbre de codage.

Codage de l'information

Exemple 7

L'arbre de codage correspondant au code instantané de l'exemple 3 ($\{0, 10, 11\}$) est



Par convention, les feuilles sont étiquetées par le symbole-source, dont le mot de code est le chemin à partir de la racine.

Codage de l'information

Définition du Code Complet

On dit que le code est un ***code complet**** lorsqu'il n'y a pas de feuille vide dans l'arbre de codage n -aire correspondant.

Codage de l'information

Inégalité du Kraft

Il existe un code instantané D -aire de N mots de code et dont les longueurs des mots de code sont les entiers positifs l_1, l_2, \dots, l_N si et seulement si:

$$\sum_{i=1}^N D^{-l_i} \leq 1$$

Lorsque l'égalité se réalise dans l'inégalité ci-dessus, le code instantané correspondant est complet.

Codage de l'information

Exemple 8

Pour le code instantané binaire complet ($D = 2$) de l'exemple 3 ($\{0, 10, 11\}$), la somme : $\sum_{i=1}^N D^{-l_i}$ est $2^{-1} + 2^{-2} + 2^{-2}$ c'est-à-dire $\frac{1}{2} + \frac{1}{4} + \frac{1}{4}$ qui est effectivement égale à 1.

De même, l'inégalité de Kraft nous apprend qu'il existe au moins un code instantané ternaire dont les longueurs de mots de code sont 1, 2, 2 et 4. En effet,

$$3^{-1} + 3^{-2} + 3^{-2} + 3^{-4} = \frac{46}{81} \cong 0.57 < 1$$

Un tel code ne serait pas complet.

Codage de l'information

Exemple 9

Existe-t-il un code binaire instantané avec des longueurs de mots de code:

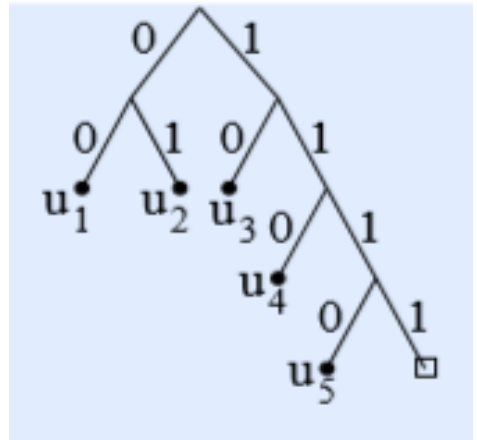
$$l_1 = 2, l_2 = 2, l_3 = 2, l_4 = 3 \text{ et } l_5 = 4 ?$$

La réponse est «**oui**», puisque

$$\sum_{i=1}^5 2^{-l_i} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16} < 1$$

L'exemple ci-dessous constitue un tel code:

| U | u_1 | u_2 | u_3 | u_4 | u_5 |
|-----|-------|-------|-------|-------|-------|
| Z | 00 | 01 | 10 | 110 | 1110 |



Moodle

| username | password | firstname | lastname | email |
|-----------------|----------|----------------|-------------|--|
| abed2016 | qj78z8m8 | Besma | ABED | besma.abed@univ-batna.dz |
| belkacem2016 | awj2522g | Widad | BELKACEM | widad.belkacem@univ-batna.dz |
| benaiacha2016 | 4q3j5x2h | Asma | BENAICHA | asma.benaicha@univ-batna.dz |
| boumekhleb2016 | wiz697q2 | Aida | BOUMEKHLEB | aida.boumekhleb@univ-batna.dz |
| bouras2016 | h5bm7z26 | Kawther | BOURAS | kawther.bouras@univ-batna.dz |
| boussenane2016 | i49ey4w3 | Aymen | BOUSSENANE | aymen.boussenane@univ-batna.dz |
| bouteghrine2016 | 66c6z2vs | Younes | BOUTEGHRINE | younes.bouteghrine@univ-batna.dz |
| chebah2016 | 29fr4vt9 | Ala eddine | CHEBAH | alaeddine.chebah@univ-batna.dz |
| chouimet2016 | 4n6xe7d3 | Rania | CHOUIMET | rania.chouimet@univ-batna.dz |
| dahmani2016 | h72xh28y | Bilal | DAHMANI | bilal.dahmani@univ-batna.dz |
| djoualla2016 | n6d7ar57 | Souhila | DJOUALLA | souhila.djoualla@univ-batna.dz |
| dziri2016 | 54ce25ki | Borhane eddine | DZIRI | borhaneeddine.dziri@univ-batna.dz |
| hocine2016 | 325bdea5 | Mohamed ilyes | HOCINE | mohamedilyes.hocine@univ-batna.dz |
| houmi2016 | unvk3898 | Ismail | HOUMI | ismail.houmi@univ-batna.dz |
| krouchi2016 | b2d8d4d2 | Mohamed | KROUCHI | mohamed.krouchi@univ-batna.dz |

Moodle

| | | | | |
|-------------------|----------|---------------------|---------------|--|
| lakhdarfriha2016 | p9pb3p78 | Achraf | LAKHDAR FRIHA | achraf.lakhdar_friha@univ-batna.dz |
| lemmouchi2016 | 5e8i82uy | Badr eddine | LEMMOUCHI | badreddine.lemmouchi@univ-batna.dz |
| m'haimedatt2016 | 4w28ugn9 | Abdallahi | M'HAIMEDATT | abdallahi.mhaimedatt@univ-batna.dz |
| mekseh2016 | 72s8mr7s | Imene | MEKSEH | imene.mekseh@univ-batna.dz |
| merazka2016 | 7iyv3w28 | Houda | MERAZKA | houda.merazka@univ-batna.dz |
| imane.merazka2016 | f57b25br | Imane | MERAZKA | imane.merazka@univ-batna.dz |
| nacercherif2016 | 59uys88n | Mohammed | NACER CHERIF | mohammed.nacercherif@univ-batna.dz |
| nya2016 | d49d2a6s | Turkiya | NYA | turkiya.nya@univ-batna.dz |
| ouazene2016 | t4v396hg | Housseayne | OUAZENE | housseayne.ouazene@univ-batna.dz |
| kouchen2016 | 4h7tw75x | Khalil fadhel allah | OUCHEN | khalil.ouchen@univ-batna.dz |
| aouchen2016 | fmg6f277 | Abdelheq | OUCHEN | abdelheq.ouchen@univ-batna.dz |
| oueznadji2016 | 4w95yr4u | Adam | OUENZADJI | adam.oueznadji@univ-batna.dz |
| ounissi2016 | sn5w48r6 | Dounia | OUNISSI | dounia.ounissi@univ-batna.dz |
| ounnar2016 | f2c895fa | Abir | OUNNAR | abir.ounnar@univ-batna.dz |
| oussaf2016 | k69ri64e | Monsif | OUSSAF | monsif.oussaf@univ-batna.dz |
| saoudi2016 | j5j64p7i | Souheyla | SAOUDI | souheyla.saoudi@univ-batna.dz |
| tabitourt2016 | 39w9ir2i | Elkhansa | TABITOURT | elkhansa.tabitourt@univ-batna.dz |
| zerouali2016 | 7z5hs5a6 | Amina maroua | ZEROUALI | aminamaroua.zerouali@univ-batna.dz |