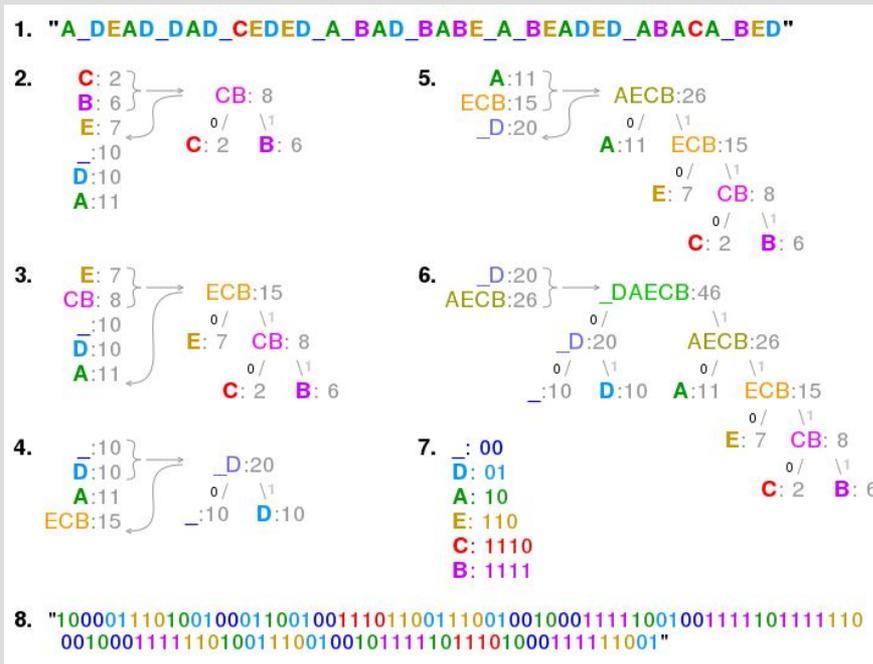


Bloc 2 : Compression avec l'algorithme d'Huffman

Groupe B2G2B

Luc Rinaudo,
Pierre Garcès,
Sébastien Dupuy,
Olivier Éloi.



Aspects pédagogiques

Pré-requis

- Numération en binaire
- Code ASCII
- Les listes en Python (partie mini-projet)

Objectifs pédagogiques

- Introduction des concepts algorithmiques dans le cas d'une situation concrète
- Introduction à la compression de données
- Extrapolation à l'application de l'algorithme d'Huffman pour compresser une chaîne de caractères

Sous-objectifs

- Les arbres binaires et les fonctions récursives en Python

Déroulé de l'activité

Partie 1 activité débranchée

“Les marmottes au sommeil léger”
(Marie Duflot-Kremer *)

Partie 2 TD

Application de la méthode de Huffman
à une chaîne de caractère

VOIR document TD

- Exposition de la problématique des marmottes au sommeil léger,
- recherche en petits groupes d'une solution optimale (affichage de la performance au tableau...*),
- partage de la meilleure solution, mise en commun,
- explicitation de l'algorithme, et application de la méthode pour tout le monde,
- partage des nouveaux résultats : discussion sur la disparité des constructions.



(*)

- <https://youtu.be/oqMx1cuw6mo>
- <https://members.loria.fr/MDuflot/files/med/marmottes.html>

Mise en situation élèves

Les marmottes au sommeil léger :

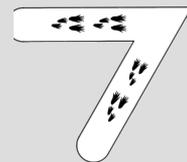
Des marmottes décident de creuser un nouveau terrier pour l'hiver, avec trois contraintes pour ne pas se déranger les unes les autres :

- A partir de l'entrée, ou à partir de l'extrémité d'un couloir, on ne peut creuser au maximum que deux couloirs, sinon la structure risque de s'effondrer.
- Une marmotte dort uniquement au fond d'une galerie qui ne donne sur rien d'autre que sa salle.
- Elles vont faire en sorte que la somme des déplacements soit la plus petite possible, pour ne pas se réveiller les unes les autres.

Construction du terrier

Chaque groupe d'élèves possède un kit de construction de terrier :

- des galeries pour accéder à la sortie du terrier ;
- des marmottes, chacune pondérée du nombre de réveils dans la période hivernale ;
- de l'explication du calcul du “nombre de déplacements” dans un terrier.



Faire un exemple du calcul pour un terrier donné au tableau.

Le revers du terrier ...

La phrase à compresser est la suivante :

BARBARA RASE BASILE LE BARBIER

							
Alicia	Bernardo	Eliane	Isidore	Louison	Roberta	Suzanna	“ ”
6	5	4	2	2	5	2	4

Écriture du dictionnaire, notion de code **préfixe**

Conclusion synthèse

C'est
complètement
OUF, Man !



- **Compression par dictionnaire** : tout symbole est associé à un mot binaire qui nécessite moins de bits que sa représentation originelle
- **Codage statistique à longueur variable** (codage entropique) : plus le symbole est fréquent, moins le mot pour le représenter est long
- **Code préfixe** : aucun mot binaire ne peut être le préfixe (début) d'un autre mot
- **Mini-projet Python** : arbres binaires et listes imbriquées
fonctions récursives (terminales ou non)

Pour aller plus loin ... ou pas

- ❖ Algorithme de Huffman : glouton, complexité $O(n \log n)$
- ❖ Démonstration d'optimalité : arbre binaire entier, lemme des fréquences faibles, lemme de propagation de l'optimalité
- ❖ Codages entropiques : Huffman, Shannon-Fano, arithmétique
- ❖ Compression par dictionnaire : Huffman, Shannon-Fano, LZ*
- ❖ Entropie de Shannon, 1er théorème de Shannon
- ❖ Comparaison moyenne Huffman / Shannon-Fano

C'est en trop !
Pie

