

Master 2 – Ingénierie Informatique (Parcours LP, SRI, LC)

Formats de Documents et Compression - Examen du 28-03-2013

Durée : 2h00

Notes manuscrites, support de cours (photocopie), et sujets de TD autorisés. Livres non autorisés. La rigueur des raisonnements, la clarté des explications, mais aussi la qualité de la présentation influera sur la note. Cinq lignes de texte au maximum pour les réponses explicatives.

Exercice 1

Une source émet le symbole **a** avec une probabilité de $0,8 = 4/5$ et le symbole **b** avec une probabilité de $0,2 = 1/5$.

1. Calculez la probabilité d'occurrence de chaque chaîne de longueur 2 sur l'alphabet $\{a, b\}$.
2. Calculez la probabilité d'occurrence de chaque chaîne de longueur 3 sur l'alphabet $\{a, b\}$.
3. Calculer un code de Huffman pour chacun de trois alphabets suivants :
 - l'alphabet $\{a, b\}$;
 - l'alphabet constitué de toutes les chaînes de longueur 2 sur l'alphabet $\{a, b\}$;
 - l'alphabet constitué de toutes les chaînes de longueur 3 sur l'alphabet $\{a, b\}$;(Attention, dans les deux derniers cas, les symboles à manipuler sont des mots!)
4. Calculer la longueur moyenne d'un mot de code dans les trois cas.
5. Ce qui arrive pour ces trois cas peut-il être généralisé à des longueurs quelconques ? Justifiez votre réponse.
6. Calculer un code de Huffman **canonique** pour le troisième alphabet.

Exercice 2

- *Bonjour Monsieur Impresario, je voudrais faire de la chanson.*
- *Il vous faut un nom de scène. Quel est votre nom complet ?*

```
01100001101101110110110011101101101111110110
      11100100000
          11000111
```

- *On va laisser tomber le nom de famille et ne garder que le prénom. Ça va marcher de tonnerre de feu !*

Quel est le nom complet de la chanteuse (prenom et nom), si on sait qu'il a été codé par Huffman adaptatif et que les codes ASCII présents codent les lettres suivantes :

```
01100001 = a
01100111 = g
01101110 = n
01101111 = o
00100000 = blanc
```

Note. Selon la convention vue en cours, dans la mise à jour de l'arbre, le nouveau noeud PEC est toujours placé comme fils droit de son père (côté 1)

Exercice 3

Une image P a été encodée par la méthode de prédiction des valeurs des pixels en utilisant les moyennes calculées sur un contexte. L'encodeur nous a donc envoyé la matrice Δ obtenue par différence de l'image d'origine P avec celle des moyennes A :

$$\Delta = P - A = \begin{matrix} & 130 & -1 & 1 & 0 \\ & 1 & 0 & 0 & \cancel{0,5} & 0 \\ & 0 & 0,5 & 0,5 & \cancel{0} & 0,5 \end{matrix}$$

1. Décodez-la en sachant que :
 - La valeur du premier pixel a été envoyée non codée.
 - En général, le contexte $C_{i,j}$ du pixel en position (i, j) est constitué des pixels en position $(i-1, j)$ et $(i, j-1)$.
 - Toutefois, le contexte des pixels de la première ligne et de la première colonne se limite uniquement aux pixels de $C_{i,j}$ déjà encodés/décodés.
2. Proposez un code binaire pour coder les valeurs de la compressée Δ d'une telle image dont les valeurs des pixels peuvent être comprises entre 129 et 135.
3. Calculez la taille de l'image compressée par le code que vous proposez. Elle devra être la plus petite possible, incluant l'espace occupé par toute information nécessaire au décodeur pour reconstruire l'image (que vous indiquerez dans les détails en spécifiant la taille qu'elle occupe).
4. Et si on sait que les valeurs des pixels dans une telle image sont comprises entre a et b , vous aurez besoin d'un code de mots de quelle longueur ? Et que devient la taille d'une image à m lignes et n colonnes ainsi codée ?
5. On veut maintenant compresser ultérieurement la séquence binaire qui code les valeurs de la matrice Δ (compression en cascade). Proposez une ou plusieurs méthodes (si plusieurs, à appliquer les unes après les autres) pour effectuer ces passes ultérieures de compression et effectuez-les sur l'image Δ . Les points obtenus seront en fonction du nombre de méthodes appliquées et du taux de compression obtenu sur Δ .