

Master 2 – Ingénierie Informatique (Parcours LP, SRI, LC)
Formats de Documents et Compression - Examen du 19-03-2012

Durée : 2h30

Notes manuscrites, support de cours (polycopié), 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 - Questions de Cours (5 points)

Vrai ou faux ? (0.5 point pour chaque réponse correcte, -0.5 points par réponse erronée)

1. Dans la méthode LZ77, utiliser un Search Buffer plus grand peut améliorer le taux de compression.
2. Dans la méthode LZ77, utiliser un Search Buffer plus grand peut améliorer la vitesse du codec.
3. Un couple de valeurs co-relies contient plus d'information redondante qu'un couple de valeurs non co-relies.
4. Pour une source qui émet les symboles :
 - a avec une probabilité de 1/2
 - b et c avec une probabilité de 1/4 chacun,il est possible de trouver un code à taille variable de longueur moyenne 1,4.
5. Le cas le plus défavorable pour la méthode de Huffman est celui où tous les symboles ont la même probabilité.
6. Dans la méthode LZ78, l'encodeur doit inclure le dictionnaire dans l'en-tête du stream codé pour que le décodeur puisse la décompresser.

Réponse à rédiger (5 lignes max - 2 points) :

7. Les vidéos sont une séquence d'images fixes affichées à intervalles très courts, ainsi donnant l'impression de mouvement. Expliquer comment vous aborderiez le problème de la compression d'un fichier vidéo, dire en particulier si et comment la méthode des transformées pourrait être adaptée à la compression de vidéos.

Exercice 2 (5 points)

Il était une fois un professeur de compression de données qui cherchait de bons exemples de phrases en langue française pour ses cours. Les hérauts du royaume avaient même diffusé en long et en large l'annonce que un point de bonus sur la note finale aurait été attribué à celui ou celle qui aurait fourni une réponse, mais aucun résultat n'était parvenu. Un après-midi, une femme habillée de manière excentrique et portant à bras un caniche, rentra sans frapper dans le bureau du professeur.

- *Bonjour, lui dit elle, j'ai entendu que vous cherchez des exemples pour vos cours. Je ne suis pas française mais je pense que mon nom peut fournir un joli exemple de Huffman adaptatif. En fait, j'adoore l'idée que pour gagner quelques points des jeunes gens doivent se casser les méninges pour proférer mon nom!*

- *Ah bon ? Et comment vous appelez-vous ?* Demanda le professeur, sans cacher son antipathie.

Pour toute réponse, elle poussa un ricanement et lui tendit une carte de visite d'excellente qualité sur laquelle était imprimé :

h0a00n10111101100 000c010101

- *Laissez-les le deviner par eux-mêmes!*
- *Mais vous croyez qu'ils vont y arriver ?*

- Et encore, j'ai été gentille! J'aurais pu vous donner la version où les lettres sont données en ASCII binaire. Une jolie séquence de 0 et 1, je vous le dis. Mais je n'ai pas voulu exagérer...

Elle se dirigea vers la porte en lui lançant sans se retourner : *Et dites à vos étudiants qu'ils devront détailler toutes les étapes de la construction de l'arbre s'ils veulent tous les points!*"

Et elle disparut ainsi, laissant le pauvre prof se demander s'il venait d'avoir une hallucination. Le chef de la cantine y était peut-être allé un peu fort avec le GMS¹ ce jour là...

Exercice 3 (4 points)

Compresser la chaîne suivante par la méthode LZ78 :

`we were there we were here`

1. Détailler les étapes de la construction du dictionnaire en supposant que celui-ci est représenté sous forme d'un arbre lexicographique. A chaque fois qu'un nouveau noeud n est créé, un numéro lui sera attribué. Ce numéro sera aussi le token du mot du dictionnaire correspondant à n .
2. Si un dictionnaire contient seulement le mot `were`, alors qu'un autre contient les deux mots `we` et `were` que faudra-t-il prévoir dans la structure de donnée pour pouvoir distinguer ces deux cas?
3. Donner une estimation aussi précise que possible du taux de compression de la chaîne codée.

Exercice 4 (6 points)

Une image P a été encodée par la méthode de prédiction des valeurs 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 :

$$\Delta = P - A = \begin{array}{cccc} 217 & -1 & 2 & 0 \\ -1 & 0 & 0 & 0.5 \\ -1 & -0.5 & 0 & 0 \end{array}$$

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 de longueur fixe pour coder les valeurs de la compressée Δ d'une telle image dont les valeurs des pixels peuvent être comprises entre 213 et 220. Votre code ne doit pas utiliser plus de 7 octets (c'est même possible de trouver un codage qui en utilise que 6) pour coder les pixels de l'image donnée et, en général, pas plus de $\frac{nm}{2} + 1$ octets pour une image à m lignes et n colonnes.
3. 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?
4. On veut maintenant utiliser plutôt un code à longueur variable pour coder les valeurs de la matrice mais on ne connaît pas une distribution de probabilité pour ces valeurs. Il faut donc proposer un modèle probabiliste (une distribution de probabilité) la plus proche possible de la réalité pour obtenir le meilleur taux de compression possible. Comment procédez-vous?

¹GMS = forme compressée de "Glutamate monosodique", un exhausteur de goût soupçonné d'avoir de mauvais effets sur la santé.