

UNIVERSITÉ DE PARIS 7, FORMATS DE DOCUMENTS ET COMPRESSION
EXAMEN DU 5 DÉCEMBRE 2005

Les documents sont autorisés. Aucun brouillon n'est accepté. La rédaction doit être claire continue (pas de renvoi à des page antérieures ou postérieures). Les exercices sont indépendants. Tout résultat doit être justifié par un minimum d'indication sur la façon dont il a été trouvé. L'étudiant ne pourra se prévaloir d'erreurs éventuelles dans les notes de cours généreusement distribuées.

EXERCICE I: HUFFMAN STATIQUE ET ADAPTATIF

On considère l'alphabet $A = \{a, b, c, d\}$ (pourquoi 2 bits sont-ils suffisants pour coder chaque symbole?).

- 1) On considère le message *baccdabdddbcdcdcdc*. Appliquer la méthode de Huffman statique. Comparer la longueur du message codé avec celle du message initiale.
- 2) Pour le message *abbccddddd*, appliquer la méthode adaptative. On dessinera l'arbre obtenu à chaque étape. Comparer la longueur du message codé avec celle du message initiale.
- 3) On considère le code de Huffman suivant sur les 8 lettres $a, b, c, d, e, f, g, h, i$.

| a | b | c | d | e | f | g | h |
|-----|-----|-----|-----|-------|-------|------|-----|
| 000 | 001 | 01 | 100 | 10100 | 10101 | 1011 | 11 |

Quelle est une distribution possible du nombre d'occurrences de lettres pour cet arbre?

- 4) Quel est l'arbre de Huffman canonique qui lui correspond (pour les symboles codés par un même nombre de lettres, on étiquettera les feuilles de la gauche vers la droite en ordre alphabétique croissant)?. Le dessiner.
- 5) On considère un texte tel que le nombre d'occurrences de chacun des 6 symboles soit donné par le tableau suivant

| a | b | c | d | e | f |
|-----|-----|-----|-----|-----|-----|
| 1 | 1 | 2 | 4 | 8 | 16 |

On veut coder, de façon optimale les symboles sur au plus 3 bits. Dire pourquoi c'est possible a priori et appliquer la méthode ad hoc vue dans le cours. Comparer les performances avec la méthode sans contrainte sur les longueurs (c'est-à-dire combien de bits par symboles sont-ils perdus?).

EXERCICE II: LZ78

On rappelle que le dictionnaire initial a une seule entrée en position 0 qui est égale à la suite vide et que le résultat de l'algorithme appliqué à un message source est une suite d'entiers.

- 1) De façon générale, peut-on voir plusieurs fois le même entier dans le message codé et si c'est le cas, combien de fois au plus?

2) Quelle est la longueur du plus long message qui peut être codé par une suite de n entiers?

3) On fixe l'alphabet à 2 lettres a et b . On considère le message $a^6b^2ab^5$. Quel est le résultat de l'application de la méthode LZ78?

EXERCICE III: LZSS

On rappelle que si l'alphabet du message A possède n lettres, le dictionnaire initial a n entrées en position $0, \dots, n-1$ contenant les n lettres dont est constitué l'alphabet.

On considère le message $a^6b^2ab^6$ sur l'alphabet $A = \{a, b\}$. Quel est le résultat de l'application de la méthode LZSS?

EXERCICE IV: PRÉDICTION

On considère l'image suivante où les entiers doivent être interprétés comme des intensités de 0 à 15, donc pouvant être codés sur un demi-octet.

| | | | | |
|----|----|----|---|----|
| 11 | 12 | 8 | 9 | 8 |
| 7 | 9 | 10 | 8 | 10 |
| 7 | 8 | 11 | 9 | 11 |
| 10 | 7 | 10 | 8 | 9 |
| 8 | 9 | 7 | 7 | 6 |

de façon générale

| | | | |
|-----|---------------|-------------|-----|
| ... | ... | ... | ... |
| ... | $a_{i-1,j-1}$ | $a_{i-1,j}$ | ... |
| ... | $a_{i,j-1}$ | $a_{i,j}$ | ... |
| ... | ... | ... | ... |

On code cette image suivant les règles suivantes.

- l'image est codée par la suite des valeurs codées de chaque position en balayant l'image du haut vers le bas et de la gauche vers la droite.
- les valeurs de la première colonne et de la première ligne sont conservées telles quelles
- toutes les autres valeurs $a_{i,j}$ sont remplacées par le codage de la valeur $a_{i,j} - \lfloor \frac{a_{i-1,j-1} + a_{i-1,j} + a_{i,j-1}}{3} \rfloor$. On codera cette différence par Huffman.

1) Donner le résultat de cette méthode sur l'image ci-dessus en remplissant le tableau 5×5 où chaque entrée comportera la donnée codée par la suite correspondante de 0 et de 1. Quel est le nombre de bits de résultant? Que manque-t-il à cette méthode pour qu'elle soit applicable?

2) Quelle est la différence maximale entre deux valeurs consécutives lorsque l'on lit les 25 valeurs du haut en bas de la gauche vers la droite? Coder l'image en donnant la valeur de l'élément $a_{1,1}$ puis en remplaçant chaque valeur par sa différence, codée par Huffman, avec la valeur de l'élément précédent dans le balayage de l'image.