

Chambre de Commerce et d'Industrie de Paris  ISBS	Unité : SP34 Examen de théorie de l'information Date : 04 février 2005	ISBS2
--	--	-------

SUJET À TRAITER – AVEC DOCUMENTS. *Tous documents autorisés*

À remettre sur une copie séparée

Remis par M. J.-F. BERCHER

**ÉNONCÉ**

Vous êtes diplômé depuis quelques années et vous occupez, entre autres tâches, de la conception et de la gestion de différents automates d'analyse biologique.

Un automate d'analyse biologique rend les résultats d'un dosage sous forme de 4 catégories notées A, B, C ou D. Lorsque le dosage est impossible pour un problème de sensibilité, le résultat rendu est X. L'automate effectue 1200 dosages par période de 24 heures (un rythme industriel !). Un de vos condisciples de l'ISBS s'est retrouvé après son diplôme et quelques péripéties dans une cambrousse éloignée et vous a contacté car une de ses connaissances médicales doit réaliser une étude statistique sur les données recueillies par les 43 automates de ce type disponibles dans le monde au cours de la dernière année, et il a pensé à vous pour lui fournir ces données. En plus, vous devez lui expédier les données par modem (Ben oui, le haut-débit ne va pas partout) et votre chef (qui est un peu spécial) n'apprécierait certainement pas que vous grêviez le budget de communications de l'entreprise. Une rapide étude statistique vous fournit les résultats suivants :

Catégories	Probabilité
A	0.3
B	0.25
C	0.25
D	0.15
X	0.15

Vous convenez avec votre camarade d'utiliser un codage entropique (dit à longueur variable, ou de Huffman) afin de limiter la durée de transmission.

1. En codant les lettres {A,B,C,D,X} sur 8 bits, combien de bits (ou d'octets, ou de kilo-octets, ou de mega-octets) devez vous utiliser pour stocker les résultats d'un automate sur un jour, un an et sur un an pour les 43 automates ?
2. Une rapide réflexion vous indique qu'il suffirait en fait de 3 bits pour coder les différents symboles. Pourquoi donc ? Combien d'octets utiliserez vous alors pour encoder la totalité des résultats fournis en un an par les 43 automates ?
3. Vous entreprenez quand même un codage de Huffman. Quelle est la quantité d'information moyenne fournie par l'automate ?
4. Quelle est la limite théorique à la longueur moyenne des mots-codes associés aux messages ? Combien pensez vous alors pouvoir économiser (en bits) par rapport au cas du codage sur 3 bits ?
5. Établissez un code de Huffman permettant d'associer un mot binaire à chacun des symboles. Vous utiliserez les conventions suivantes : lorsque vous regrouperez deux probabilités, vous associerez un 0 au symbole correspondant à la probabilité la plus élevée et un 1 au symbole de probabilité la plus faible ; en cas d'égalité des probabilités, vous classerez le nouveau symbole en « haut ».
6. Quelle est la longueur moyenne des messages ? Le code obtenu est-il déchiffrable ? Quel est finalement le nombre moyen de bits que vous devrez utiliser pour encoder la totalité des résultats ?
7. Transformez le message ABBXACDCAAXCB

8. Vous devez expliquer à votre camarade de promotion comment il devra s'y prendre pour décoder les données, vu qu'il semble avoir un peu oublié (et vous nous faites partager cette explication).
9. Quel(s) autre(s) algorithme(s) sans perte auriez vous pu utiliser (Citez en au moins 1) ?

Votre camarade de promotion, enchanté de vos travaux, vous passe un coup de téléphone de son lointain exil (mais il y est au chaud, le bougre) et vous interroge l'air de rien sur les meilleures manières de compresser des images — en effet, vous l'apprendrez plus tard, il est payé fort cher comme expert en biosciences et une partie de sa mission consiste à sauvegarder des photographies de la biosphère locale (qui est tout-à-fait particulière et très rare, on s'en doute). Comme il était un peu "distrait" pendant les cours de l'ISBS, mais qu'il est très habile, il arrive à vous convaincre de rédiger par écrit les réponses à quelques interrogations (et vous nous faites partager vos réponses...) :

1. Quels sont les critères à prendre en compte pour choisir ou construire une méthode de compression d'images ?
2. Un codage RLC permet il d'obtenir une compression ? Dans l'affirmative, dans quelles conditions ?
3. Quelles sont les différences entre TFD et DCT ? Quels sont les avantages de la DCT par rapport à la transformée de Fourier ?
4. Donnez les grands principes de la compression JPEG avec pertes (processus de fond). Quel paramètre permet de régler le taux de compression ? Pourquoi doit on quantifier les coefficients DCT ?
5. Si on applique une compression JPEG à une image 256 couleurs (3 images primaires tel que le nombre de couleurs total soit de 256), l'image décompressée possèdera t'elle
  - plus de couleurs,
  - le même nombre de couleurs,
  - moins de couleurs ?

Surpris par votre patron pendu au téléphone à évoquer le bon vieux temps, vous vous prenez un savon et en profitez pour accepter la proposition de votre camarade qui vous propose de le rejoindre dans sa savane (où il compte bien sûr exploiter vos connaissances au maximum...). — FIN

*Deux remarques — Il n'y a pas dans le texte qui précède de conseil de choix de carrière, et je n'ai rien de particulier contre les automates biologiques...*