TD d'algorithmique numéro 7
Olivier Bodini et Natacha Portier
lundi 5 novembre 2001

Question 1 $\star$ codage de Huffman $\star$

Soit Σ un alphabet, c'est à dire un ensemble fini, que l'on suppose de cardinal au moins deux. Ses éléments sont appelés des lettres. Un codage binaire est une application injective de Σ dans l'ensemble des suites finies de 0 et de 1. Il s'étend de manière naturelle par concaténation à une application définie sur l'ensemble Σ^* des mots de Σ, c'est à dire des suites finies de Σ. Un codage est **de longueur fixe** si tous les caractères sont codés par un mot binaire de même longueur. Il est **préfixe** si aucun mot du code n'est préfixe d'un autre mot du code.

Question 1.1

Le décodage d'un codage de longueur fixe est unique. Montrer qu'il en est de même pour un codage préfixe.

Question 1.2

Représenter un codage préfixe par un arbre binaire dont les feuilles sont les lettres de l'alphabet.

On considère un texte dans lequel chaque caractère c apparaît avec une fréquence $f(c)$ non nulle. À chaque codage préfixe de ce texte, représenté par un arbre T, est associé un *coût*, défini par

$$B(T) = \sum_{c \in \Sigma} f(c) d_T(c)\,,$$

où $d_T(c)$ est la taille du mot binaire codant c. Un codage préfixe représenté par l'arbre T est optimal si pour ce texte il minimise la fonction B. Montrer qu'à un codage préfixe optimal correspond un arbre binaire complet, i.e. dans lequel tout noeud interne a deux fils. Montrer qu'un tel arbre a $|\Sigma|$ feuilles et a $|\Sigma| - 1$ nœuds internes.

Question 1.3

Montrer qu'il existe un code préfixe optimal pour lequel les deux caractères de plus faible fréquence sont frères, i.e. leurs codes sont de même longueur et ne diffèrent que par le dernier bit (*propriété du choix glouton*).

Soit T l'arbre d'un codage binaire optimal pour la fréquence f sur l'alphabet Σ. Soient x et y deux lettres soeurs et soit z une lettre fraîche avec laquelle on étiquette leur mère. Alors, en considérant z comme un caractère de fréquence $f(z) = f(x) + f(y)$ montrer que l'arbre $T' = T - \{x, y\}$ représente un codage préfixe optimal pour l'alphabet $\Sigma' = \Sigma - \{x, y\} \cup \{z\}$ (*propriété de sous-structure optimale*).

Question 1.4

Donner un algorithme pour trouver un codage optimal. Trouver le codage minimal pour $\Sigma = \{a, b, c, d, e, g\}$ et $f(a) = 45$, $f(b) = 13$, $f(c) = 12$, $f(d) = 16$, $f(e) = 9$ et $f(g) = 5$.

Question 2

$\star$ chaînes d'une permutation $\star$

On considère une permutation $p = (p[1], \ldots, p[n])$ de $[1, \ldots, n]$. On pose $p_i = (i, p[i]) \in \mathbb{N}^2$; on dit que p_i *domine* p_j si $i \geq j$ et $p[i] \geq p[j]$. On écrit cela $p_i \geq p_j$, et on note $p_i > p_j$ pour $p_i \geq p_j$ et $p_i \neq p_j$.

On vérifie aisément que la relation $\geq$ est un ordre partiel.

Une *chaîne* est une suite $p_{i_1}, \ldots, p_{i_k}$ telle que $p_{i_1} > \cdots > p_{i_k}$. On définit la *hauteur* de p_i comme la longueur maximale d'une chaîne dans l'ensemble des p_j tels que $p_i \geq p_j$. Soit S_h l'ensemble des points de hauteur h.

Question 2.1

On fixe i, et on note h la hauteur de p_i. Montrer que le point de S_{h-1} le plus à droite parmi ceux qui sont à gauche de p_i est dominé par p_i (les p_i sont vus comme des points du plan).

Question 2.2

Proposer un algorithme pour calculer les hauteurs des éléments d'un ensemble de n points.

Comment exhibe-t-on une chaîne de longueur maximale?