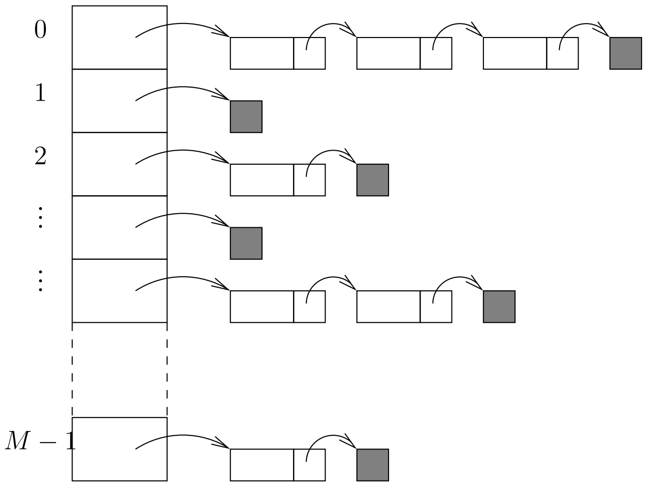
L3-INFO Algorithmique

Université Joseph Fourier
GRENOBLE
UFR IM²AG

ALGO5 : Travaux dirigés, séance 5
Hachage

1 Table de hachage

On considère le problème de hachage fermé dans une table de taille M. La fonction de hachage est notée h et le nombre de clés à "hacher" est N. Dans ce TD on considère un hachage fermé. c'ést à dire que la résolution de collision se fait par chaînage externe à la table.



On suppose que la fonction de hachage h est uniforme. Ceci se modélise en associant à chaque clé x de l'éspace des clés une variable aléatoire U_x de loi uniforme à valeurs dans $\{0, \cdots, M-1\}$. On suppose que ces variables aléatoires sont indépendantes entre elles. Pour notre problème, on notera les N variables aléatoires $\{U_1, U_2, \cdots, U_N\}$.

L'état de remplissage de la table est modélisé par un vecteur de dimension M : $X = [X_0, \cdots, X_{M-1}]$, où X_i représente le nombre de clés hachées sur la case i.

Question 1.1 : Calcul des X_i

Exprimer les X_i en fonction des U_i. Pour i donné calculer la loi de X_i.

Question 1.2 : Recherche infructueuse

Calculer le coût moyen d'une recherche d'une clé qui n'ést pas dans la table. On notera $\alpha = \frac{N}{M}$ le taux de remplissage de la table.

Question 1.3 : Recherche d'un élément de la table

Calculer le coût moyen d'une recherche d'une clé que l'on sait dans la table.

Question 1.4 : Insertion et suppression d'un élément de la table

Calculer le coût moyen d'une insertion d'une nouvelle clé dans la table. Calculer le coût moyen d'une suppression d'une clé de la table.

Hachage

Question 1.5 : Taux de cases vides

Calculer la probabilité qu'une case soit vide, donner un équivalent en fonction de α pour N et M grands. En déduire le nombre moyen de cases vides dans la table.

Question 1.6 : Approximation

On pourra vérifier (à la maison) que pour n pas trop grand et M et N grands

$$\mathbb{P}(X_i = n) \simeq e^{-\alpha}\frac{\alpha^k}{k!}.$$

Approximation par une loi de Poisson.

En utilisant l'inégalité

$$e^\alpha \leqslant 1 + \frac{\alpha}{1!} + \frac{\alpha^2}{2!} + \cdots + \frac{\alpha^k}{k!} + \frac{\alpha^{k+1}}{(k+1)!}e^\alpha,$$

montrer que

$$\mathbb{P}(X_i > k) \leqslant \frac{\alpha^{k+1}}{(k+1)!}$$

En déduire que

$$\mathbb{P}(X_i > b\alpha) \leqslant (\frac{e}{b})^{b\alpha}.$$

On pourra utiliser la formule de Stirling $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

Comment interpréter cette quantité ? Pourquoi peut-on parler de bon comportement de la recherche en table ? Calculer pour différentes valeurs de b et de α cette probabilité.

Panini

Votre petit cousin collectionne les vignettes autocollantes des footballeurs célèbres. Il achète au marchand de journaux des pochettes de 10 vignettes à un prix de 2 euros. L'album de la collection complète contient $M = 300$ places pour les vignettes autocollantes (prix d'achat de l'album 4 euros). Il souhaite savoir combien de pochettes il devra acheter pour avoir toute la collection. Saurez-vous lui répondre ? On simplifie le modèle en supposant que l'on achète les vignettes une par une et on note T le nombre total de vignettes achetées pour avoir toute la collection. A la fin, il restera plein de doubles ! On supposera également que l'éditeur des vignettes est honnête et qu'il imprime les vignettes dans les mêmes proportions (il n'y a pas de vignettes plus rares que d'autres). Lorsque la collection comporte déjà $i - 1$ vignettes, on note Y_i le nombre de vignettes à acheter pour avoir une vignette qui nést pas dans la collection.

Question 2.1 :

Donner, en la justifiant, la loi de Y_i. Les variables Y_i sont-elles indépendantes.

Question 2.2 :

Exprimer T en fonction des Y_i. En déduire $\mathbb{E}T$ et $VarT$. Commenter votre résultat et répondre à votre petit cousin.

Question 2.3 :

Quel est le rapport entre les panini et les tables de hachage ?