TD n^{o} 8 - Tables de hachage

Exercice 1. *Un exemple*

On dispose d'une table de hachage de taille 9 dont les collisions sont résolues par liste chaînée, et dont la fonction de hachage est $h(k) = k \mod 9$.

✎ Exécutez l'algorithme d'insertion dans la table des clés suivantes : $5, 28, 19, 15, 20, 33, 12, 17, 10$.

Exercice 2. *Sondage linéaire*

On va maintenant étudier une méthode pour résoudre les collisions dans les tables de hachage, sans utiliser de listes chaînées.
Ici, on met directement les entrées de la table dans les cases du tableau (une seule entrée par case). Une première conséquence immédiate est qu'il n'est pas possible de mettre plus d'entrées qu'il y a de cases dans le tableau (alors qu'avec les listes chaînées, ce n'était pas un gros problème).
Par ailleurs, lorsque l'on essaie d'insérer une nouvelle entrée, si la case où l'on veut la mettre est déjà occupée, il faut essayer une autre case, et ainsi de suite jusqu'à en trouver une vide. On dit alors que l'on *sonde* la table pour trouver une case libre. Il faut alors modifier la fonction de hachage, qui prend maintenant un argument i de plus, qui est le nombre de tentatives infructueuses que l'on a fait jusque là :

$$h : \begin{cases} \mathbb{N} \times \mathbb{N} & \rightarrow & [\![0, m-1]\!] \\ (k, i) & \mapsto & h(k, i) \end{cases}$$

On considère ici la fonction de hachage

$$h : \begin{cases} \mathbb{N} \times \mathbb{N} & \rightarrow & [\![0, m-1]\!] \\ (k, i) & \mapsto & (h'(k) + i) \mod m \end{cases}$$

où h' est une fonction de hachage « simple » (c'est-à-dire comme celles que l'on utilisait avec les listes chaînées).

1. Exécutez à la main l'algorithme d'insertion dans une table de taille $m = 9$ des valeurs 5, 28, 19, 15, 20, 33, 12, 17 et 10 en utilisant la fonction $h'(k) = k \mod 9$.

2. Écrivez les fonctions permettant d'insérer et de rechercher une clé dans une table de hachage de taille m à adressage ouvert avec sondage linéaire (ce que l'on vient de décrire). On suppose que la fonction h' est déjà définie sous le nom `h1`.

3. Que se passe-t-il si l'on supprime certaines clés de la table ? Comment peut-on résoudre ce problème ?

Le sondage linéaire présente cependant un gros inconvénient. Même si la fonction de hachage répartit les clés uniformément dans la table, la résolution des collisions a tendance à former des « paquets » de valeurs...

4. Dans une table de taille m, quelle est la probabilité qu'une clé soit placée dans une case vide précédée d'une autre case vide ? Quelle est la probabilité qu'une clé soit placée dans une case vide précédée de k cases pleines ?

5. Pourquoi le phénomène de « paquets » est-il mauvais pour l'efficacité des tables de hachage ?

Exercice 3. *Sondage quadratique*

Une première solution pour résoudre le problème des « paquets » est de modifier la méthode de sondage pour que les cases sondées soient de plus en plus éloignées. C'est ce que font les fonctions de hachage par sondage quadratique :

$$h : \begin{cases} \mathbb{N} \times \mathbb{N} & \rightarrow & [\![0, m-1]\!] \\ (k, i) & \mapsto & (h'(k) + c_1 i + c_2 i^2) \mod m \end{cases}$$

Bien que le phénomène de groupage soit moins flagrant avec le sondage quadratique, si deux clés k_1 et k_2 ont la même valeur de hachage initiale ($h(k_1, 0) = h(k_2, 0)$) alors la séquence de sondage pour ces deux clés sera identique (on cherchera à les insérer dans les mêmes cases).

1. Pourquoi est-il mauvais de ne pas utiliser beaucoup de séquences de sondage différentes ?

Exercice 4. *Double hachage*

Le double hachage est l'une des meilleures méthodes connues pour l'adressage ouvert. Il utilise une fonction de hachage de la forme

$$h : \begin{cases} \mathbb{N} \times \mathbb{N} & \rightarrow & [\![0, m-1]\!] \\ (k, i) & \mapsto & (h_1(k) + i h_2(k)) \mod m \end{cases}$$

où h_1 et h_2 sont des fonctions de hachage auxiliaires « simples ».

1. Insérez les clés de la questions 1 de l'exercice 1 dans une table de taille $m = 13$ en utilisant le double hachage $h(k, i) = h_1(k) + i h_2(k)$ avec $h_1(k) = k \mod 13$ et $h_2(k) = 1 + (k \mod 12)$.

2. Que se passe-t-il si $h_2(k)$ et m ne sont pas premiers entre eux ? En quoi est-ce gênant ?

3. Combien de séquences de sondage différentes peut-on obtenir par la méthode du double hachage ?

Exercice 5. *Collisions*

On dispose d'une table de hachage de taille m dont les collisions sont résolues par liste chaînée. On suppose que l'on a effectué un hachage uniforme de n clés dans la table par une fonction h.

1. Quel est le nombre attendu de collisions, c'est-à-dire l'espérance du cardinal de $\{(x, y) \mid h(x) = h(y)\}$?

2. Quel est le nombre moyen de cases vides dans le tableau, c'est-à-dire le cardinal moyen de

$$\{i \in [\![0, m-1]\!] \mid \forall k \in K, h(k) \neq i\}$$

où K est l'ensemble des clés qui ont été insérées dans la table ?