

## TDP N° 4

### Structures de Données Probabilistes

#### Partie 01 : TD

#### Exercice 1: Table de hachage, gestion des collisions par adressage ouvert

On utilise une fonction  $h_1$  à 2 arguments  $h_1(c,i)$  : tel que  $i$  = nbr d'essais de placement déjà effectués.

On considère la fonction de hachage :

$$h_1(c,i) = (h'(c) + i*a) \bmod(m) \text{ avec } h'(c) = c \bmod(m) \text{ tels que } m = 11 \text{ et } a = 1$$

1. Représentez l'état de la table de hachage **T1** après le calcul de la fonction  $h_1(c,i)$  appliqué à  $c = 33; 37; 6; 99; 16; 26; 50; 70; 4$

2. Refaire la même question pour la table de hachage **T2** après le calcul de la fonction  $h_2(c,i)$  appliqué à  $c = 33; 37; 6; 99; 16; 26; 50; 70; 4$  avec

$$h_2(c,i) = (h''(c) + i*a) \bmod(m) \text{ tels que } h''(c) = 3*c \bmod(m) \text{ et } m = 11, \underline{a=2}$$

3. Refaire la même question pour la table de hachage **T3** après le calcul de la fonction  $h_3(c,i)$  appliqué à  $c = 33; 37; 6; 99; 16; 26; 50; 70; 4$  avec

$$h_3(c,i) = (L(c) + i^2) \bmod(m) \text{ tels que } L(c) = 3*c + 4 \bmod(m) \text{ et } \underline{m = 11},$$

4. Refaire la même question pour la table de hachage **T4** après le calcul de la fonction  $h_4(c,i)$  appliqué à  $c = 33; 37; 6; 99; 16; 26; 50; 70; 4$  avec

$$h_4(c,i) = (R(c) + i*S(c)) \bmod(m) \text{ tels que } R(c) = 3*c + 4 ; S(c) = 2*c + 5 \text{ et } \underline{m = 11},$$

Table T1		Table T2		Table T3		Table T4	
0		0		0		0	
1		1		1		1	
2		2		2		2	
3		3		3		3	
4		4		4		4	
5		5		5		5	
6		6		6		6	
7		7		7		7	
8		8		8		8	
9		9		9		9	
10		10		10		10	

### Exercice 2: Bloom Filter

Soit la Table B qui représente un Bloom Filter vide de 13 bits.

La première ligne représente les indices (positions) et la deuxième ligne représentent les bits. Tous les bits doivent être initialisés à 0.

Soient les deux fonctions de hachage liées au bloom filter :

- $H1(x) = 2x + 1 \bmod(13)$
- $H2(x) = 5x + 3 \bmod(13)$

En utilisant les fonctions de hachage H1 et H2, réaliser l'insertion dans le Bloom Filter des éléments suivants : 17, 87, 25, 7, 5, 11, 4, 96, 21, 48

**Table B. Bloom filter de taille 13**

Indice	0	1	2	3	4	5	6	7	8	9	10	11	12
Valeur													

### Partie 02 : TP

**Exercice 1:** Table de Hachage (TH) - gestion des collisions par chaînage.

1. Ecrire les fonctions permettant d'ajouter, rechercher ou supprimer un élément ( nombre entier ) dans une table de hachage de taille m en utilisant la gestion de collision par chaînage et des structures de données de votre choix ( liste, dictionnaire, combinaison liste & dictionnaire, ... ). Prendre la FH  $H(c) = c[m]$ ,
2. Prendre  $m = 100, 1000, 10000$  puis 100000 et à chaque fois  $n = 2*m$ . Générer aléatoirement n entiers puis insérer les dans la TH.

**Exercice 2:** Table de hachage, gestion des collisions par adressage ouvert

On utilise une fonction h à 2 arguments  $h(c,i,a)$  : tel que i = nbr d'essais de placement déjà effectués et le nombre **a** représente le pas.

Ajouter un élément de clé c à la table revient alors à regarder si la cellule d'indice  $h(c,0,a)$  est vide. Si c'est le cas, on y met l'élément, sinon on essaie la cellule d'indice  $h(c,1,a)$ , puis  $h(c,2,a)$ , ...

On considère la fonction de hachage :

$$h(c,i,a) = (h'(c) + i*a) \bmod(m) \text{ avec } h'(c) = c \bmod(m) \text{ avec } m = 11.$$

1. Proposer deux autres méthodes de sondage pour gérer les collisions :
  - Sondage quadratiques
  - Double hachage
2. Ecrire les fonctions permettant d'ajouter, rechercher ou supprimer un élément dans une table de hachage pour chaque manière de sondage.
3. Prendre  $m \sim 10, 100, 1000$  puis  $10000$  (vous pouvez choisir  $m$  un nombre premier) et à chaque fois  $n = 0,7*m$ .

Générer aléatoirement  $n$  entier entre  $0$  et  $m-1$  puis insérer les dans la THen utilisant : sondage linéaire, quadratique et double hachage. Comparer et ploter les temps d'exécutions de chaque méthode.

**Exercice 3:** Implémenter le filtre de Bloom vu au cours, en prenant les paramètres  $(n, m, k) = (5 \cdot 10^3, 25 \cdot 10^3, 3)$ . Les éléments traités seront des entiers générés aléatoirement et à vous de proposer les  $k$  Fonctions de Hachage.

Prendre  $k = 2, 3$  et  $k = 4$ , puis comparer le pourcentage des faux positifs pour  $k=2, 3$  et  $4$ .

**Exercice 4:** Implémentation d'un CMS

En utilisant les notations du cours, implémenter un CMS en choisissant les paramètres adéquats de sorte à montrer l'intérêt d'augmenter le nombre de fonctions de hachages pour améliorer la précision.