

## TD 4 : Fonctions de hachage

---

**Exercice 1.***Un hachage sans collision*

Une fonction de hachage  $h : U \rightarrow \{0, \dots, m-1\}$  est *sans collision* pour un ensemble  $X \subset U$  si pour tout  $x, y \in X$ ,  $h(x) \neq h(y)$ . Dans cet exercice, on suppose  $X$  fixé.

1. Donner une condition nécessaire et suffisante sur  $X$  pour qu'il existe une fonction de hachage sans collision pour  $X$ .
2. Supposons qu'on ait choisi une fonction  $h$  aléatoire. Exprimer l'espérance du nombre de collisions pour  $X$  en fonction de  $m$  et  $n = |X|$ .
3. Quelle est la probabilité qu'une fonction aléatoire  $h$  soit sans collision pour  $X$ .
4. Supposons qu'on cherche une fonction sans collision pour  $X$  en tirant des fonctions aléatoires tant qu'on en a pas trouvé une qui convienne. Quelle est l'espérance du nombre de tirages nécessaires ?

**Exercice 2.***Filtres de Bloom*

On s'intéresse dans cet exercice à une structure de données qui permet de stocker de manière très compressée un ensemble (statique, c'est-à-dire duquel on ne supprime jamais d'élément). La contrepartie est la présence de faux-positifs : la structure de données répond parfois que  $x$  appartient à l'ensemble alors que ça n'est pas le cas. Son utilisation en pratique vient en appui d'une *vraie* structure de donnée, pour fournir un pré-test d'appartenance très rapide<sup>1</sup>.

Un filtre de Bloom pour un ensemble de taille  $n$  est donné par un entier  $m$  (la taille de la représentation) et  $k$  fonctions de hachage  $h_1, \dots, h_k$  indépendantes. Un ensemble  $X$  est représenté par un mot booléen  $w$  de taille  $m$ . L'ensemble vide est représenté par le mot  $0 \dots 0$ . Pour insérer un nouvel élément  $x$ , on passe à 1 les  $k$  bits de  $w$  d'indices  $h_1(x), \dots, h_k(x)$ . Un bit peut être mis plusieurs fois à 1. Pour tester si un élément  $y$  appartient à  $X$ , on vérifie si  $w_{h_j(y)}$  vaut 1 pour  $1 \leq j \leq k$  : si c'est le cas, on répond « oui » et sinon on répond « non ».

Dans la suite, on suppose qu'on a construit la représentation  $w$  d'un ensemble  $X$  de taille  $n$ . On se place dans le modèle aléatoire pour les fonctions de hachage.

1. Laquelle des deux réponses de l'algorithme de recherche est toujours exacte ?
2. Montrer que le  $i$ -ème bit  $w_i$  de  $w$  vaut 1 si et seulement s'il existe  $x \in X$  et  $j$  tels que  $h_j(x) = i$ .
3. Quelle est la probabilité  $p$  que le  $i$ -ème bit de  $w$  soit égal à 0 ? *On rappelle qu'on se place dans le modèle aléatoire, et que la probabilité dépend du choix des fonctions de hachage.*

On fait maintenant l'hypothèse qu'une fraction  $p$  des bits de  $w$  sont à 0.

4. Pourquoi cette hypothèse ne découle pas de la question précédente ?
5. Soit  $y \notin X$ . Quelle est la probabilité d'obtenir un faux-positif, c'est-à-dire que l'algorithme de recherche réponde « oui » sur l'entrée  $y$  ?
6. Montrer qu'en prenant  $k = m \ln 2/n$ , cette probabilité est exponentiellement petite. *On pourra utiliser, entre autres, que  $1 - x \geq e^{-2x}$  pour  $x \leq 1/2$ .*

---

<sup>1</sup>. Voir [https://en.wikipedia.org/wiki/Bloom\\_filter#Examples](https://en.wikipedia.org/wiki/Bloom_filter#Examples) pour de nombreux exemples d'utilisation de ces objets en pratique.