Hachage par chaînage

Référence : Cormen

Introduction. On considère un cache où les collisions sont résolus par chaînage, et on se place dans l'hypothèse de **hachage uniforme simple** : chaque élément a la même chance d'être haché vers l'une des quelconque alvéoles indépendamment des endroits où les autres éléments sont allés.

Pour une table de hachage à m alvéoles et n éléments, on note $\alpha = n/m$ le facteur de remplissage.

Théorème 1. *Dans une table de hachage pour laquelle les collisions sont résolues par chaînage, une recherche infructueuse prend un temps moyen* $\Theta(1 + \alpha)$*, sous l'hypothèse d'un hachage simple uniforme.*

Démonstration. On note $\mathbf{Rech}^-(T, k)$ la variable aléatoire qui compte le nombre de comparaisons dans une recherche infructueuse d'un élément k dans un tableau T à n éléments.

$$\mathbb{E}(\mathbf{Rech}^-(T, k)) = \sum_{i=1}^{m} Pr[h(k) = i]\lambda_i$$

où λ_i est la taille de l'alvéole i de T.

Par hypothèse de hachage uniforme simple,

$$\mathbb{E}(\mathbf{Rech}^-(T, k)) = \frac{1}{m}\sum_{i=1}^{m} \lambda_i = \alpha$$

Maintenant, on note $\mathbf{Rech}^-(n, m)$ le nombre de comparaison dans un tableau déjà rempli avec n éléments $k_1, ..., k_n$. Par hypothèse de hachage uniforme simple, le tableau est tiré uniformément parmi l'ensemble des tableaux à m alvéoles et n éléments. Ainsi, par linéarité de l'espérance,

$$\mathbb{E}(\mathbf{Rech}^-(n, m)) = \frac{1}{m^n}\sum_{T} \mathbb{E}(\mathbf{Rech}^-(T, k))$$

et donc $\mathbb{E}(\mathbf{Rech}^-(n, m)) = \alpha$. □

Théorème 2. *Dans une table de hachage pour laquelle les collisions sont résolues par chaînage, une recherche infructueuse prend un temps moyen* $\Theta(1 + \alpha)$*, sous l'hypothèse d'un hachage simple uniforme.*

Démonstration. L'idée est de prendre aléatoirement un élément qui est déjà dans le tableau, et compter le nombre de comparaisons nécessaires. Le nombre d'éléments examinés pour un éléments x est 1 plus le nombre d'éléments suivants qui ont été haché au même endroit.

On suppose que l'on a inséré les éléments $x_1, ..., x_n$ dans la table de hachage, chacun de clé $k_1, ..., k_n$. On note $\mathbf{Rech}^+(x_i)$ le nombre de comparaisons nécessaires pour trouver x_i. De plus, on note $X_{ij} = \mathbf{1}\{h(k_i) = h(k_j)\}$. Par hypothèse de hachage simple uniforme, $\mathbb{E}(X_{ij}) = 1/m$. Ainsi,

$$\begin{aligned}
\mathbb{E}(\mathbf{Rech}^+(n,m)) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\mathbf{Rech}^+(x_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(1+\sum_{j=i+1}^{n}X_{ij}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(1+\sum_{j=i+1}^{n}\mathbb{E}(X_{ij})\right) \\
&= 1+\frac{1}{n}\sum_{1\leqslant i<j\leqslant n}\frac{1}{m} \\
&= 1+\frac{1}{n}\frac{n(n-1)}{2m} \\
&= 1+\frac{\alpha}{2}-\frac{1}{2m}
\end{aligned}$$

□