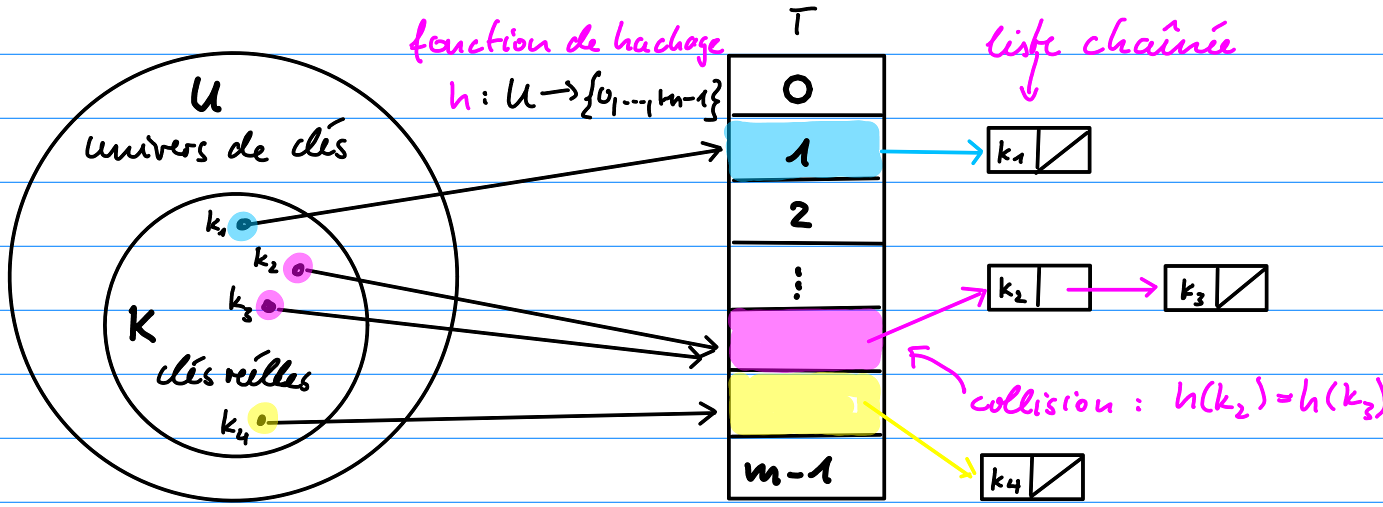
TABLES DE HACHAGE (avec chaînage)

table de hachage : structure de données permettant d'implémenter efficacement des dictionnaires



opérations (résolution des collisions par chaînage)

- **RECHERCHER** (T, k) x in $T[h(\text{clé}(x))]$ tq $\text{clé}(x) = k$
- **INSÉRER** (T, x) : $T[h(\text{clé}(x))]$.append(x)
- **SUPPRIMER** (T, x) $T[h(\text{clé}(x))]$.remove(x)

Facteur de remplissage α d'une table de hachage avec m alvéoles qui stocke n éléments : $\alpha := \frac{n}{m}$
(c-à-d le nombre moyen d'éléments stockés dans une liste)
$n = n_0 + n_1 + \dots + n_{m-1}$, où $n_i := \text{len}(T[i])$

hypothèses

- $h(k)$ peut être calculé en temps $O(1)$
- la recherche d'un élément x dans $T[h(\text{clé}(x))]$ coûte $O(n)$ où $n = \text{len}(T[h(\text{clé}(x))])$
- chaque élément est haché vers l'une quelconque des alvéoles (hachage uniforme simple)

Théorème : Une recherche infructueuse prend un temps moyen de $\Theta(1+\alpha)$

Démo : Le temps moyen pour la recherche infructueuse d'une clé k est le temps moyen pour la poursuite de la recherche jusqu'à la fin de $T[h(k)]$, qui a une longueur moyenne $E[n_{h(k)}] = \alpha$
$\Rightarrow$ coût y compris le coût du calcul de $h(k)$: $\Theta(1+\alpha)$

$\square$

Théorème : Une recherche réussie prend en moyenne un temps $\Theta(1+\alpha)$

Démo : Le nombre d'éléments examinés lors d'une recherche fructueuse portant sur x est egal à
$1+$ le nombre d'éléments qui apparaissent avant x dans la liste $T[h(\text{clé}(x))]$
$= 1+$ le nombre d'éléments y tq $h(\text{clé}(y)) = h(\text{clé}(x))$ insérés avant x

Soient $x_1, \dots, x_n$ les éléments insérés dans la table T et $k_i = \text{clé}(x_i)$, $1 \leq i \leq k$.

Soit $X_{ij} = \begin{cases} 1 & \text{si } h(k_i) = h(k_j) \\ 0 & \text{sinon} \end{cases}$

En supposant qu'il y a hachage uniforme simple on a $Pr[X_{ij} = 1] = 1/m$, donc $E[X_{ij}] = 1/m$

Le nombre moyen d'éléments examinés dans une recherche réussie est

\# moyen d'éléments ajoutés avant x_i avec clé k_i

$$E\Big[\frac{1}{n}\sum_{i=1}^{n}\Big(1+\sum_{j=1}^{i-1} X_{ij}\Big)\Big]$$

linéarité de l'espérance

$$= \frac{1}{n}\sum_{i=1}^{n}\Big(1+\sum_{j=1}^{i-1} E[X_{ij}]\Big)$$

$1/m$

$$= \frac{1}{n}\sum_{i=1}^{n}\Big(1+\frac{i-1}{m}\Big) = 1 + \frac{1}{nm}\sum_{i=0}^{n-1} i$$

$$= 1+\frac{n-1}{2m} = 1+\frac{\alpha}{2}-\frac{\alpha}{2n} = \Theta(1+\alpha) \qquad \square$$

Donc si $n = O(m)$ alors $\alpha = O(1)$. Par conséquent, recherche, insertion et suppression prennent un temps $O(1)$

HACHAGE UNIVERSEL

Supposons qu'un ennemi choisit les clés à hacher via une fonction de hachage figée, il peut choisir n clés qui sont hachés vers la même alvéole

$\Rightarrow$ recherche en temps $\Theta(n)$ en moyenne !

<u>stratégie</u> : choisir la fonction de hachage aléatoirement (indépendamment des clés qui vont être stockées)

Def : Une collection finie $\mathcal{H}$ de fonctions $U \to \{0,1,\dots,m-1\}$ est dit universelle si pour chaque paire de clés $k, \ell \in U$ distinctes, le nombre de fonctions $h \in \mathcal{H}$ pour lesquelles $h(k) = h(\ell)$ vaut au plus $|\mathcal{H}|/m$.

(les chances de collision entre k et ℓ sont au plus $\frac{1}{m}$, si k et sont choisis aléatoirement et indépendamment)

Théorème : Soit $\mathcal{H}$ une collection universelle des fonctions de hachage. Si $h \in \mathcal{H}$ est choisie aléatoirement et la clé k n'est pas dans la table alors la longueur moyenne $E[n_{h(k)}]$ de la liste $T[h(k)]$ est d'au plus α.
Si la clé k est dans la table alors $E[n_{h(k)}] \leq 1+\alpha$.

Démo : Pour clés distinctes k, ℓ soit $X_{k\ell} = \begin{cases} 1 & \text{si } h(k)=h(\ell) \\ 0 & \text{sinon} \end{cases}$

On a $\Pr[X_{k\ell}=1] \underset{\uparrow}{\overset{\mathcal{H} \text{ est universel}}{\leq}} \frac{1}{m}$ et $E[X_{k\ell}] \leq \frac{1}{m}$

Pour toute clé k soit $Y_k = \sum_{\text{clé } \ell \neq k} X_{k\ell}$.

Donc $E[Y_k] = \sum_{\text{clé } \ell \neq k} E[X_{k\ell}] \leq \sum_{\text{clé } \ell \neq k} \frac{1}{m}$ (*)

- Si k n'est pas dans la table, alors $n_{h(k)} = Y_k$ et $|\{\ell \mid \ell \text{ est dans la table et } \ell \neq k\}| = n$.
  Donc $E[n_{h(k)}] = E[Y_k] \leq \frac{n}{m} = \alpha$

- Si k est dans la table alors k apparaît dans $T[h(k)]$
  Puisque Y_k n'inclut pas la clé k on a $n_{h(k)} = Y_k + 1$ et $|\{\ell \mid \ell \text{ est dans la table et } \ell \neq k\}| = n-1$
  Donc $E[n_{h(k)}] = E[Y_k] + 1 \leq (n-1)/m + 1 = 1 + \alpha - \frac{1}{m} < 1+\alpha$

$\square$

Selon le corollaire suivant si l'on utilise le hachage universel il est impossible à un ennemi de sélectionner une suite d'opérations qui force le temps d'exécution le plus défavorable.

Corollaire : Avec le hachage universel il faut un temps moyen $\Theta(n)$ pour traiter une suite quelconque de n opérations d'insertion, suppression et recherche contenant $O(m)$ opérations INSERER

n° d'alvéoles

Démo : appliquer le théorème précédent et utiliser la linéarité de l'espérance