Leçon 5 : Tables de hachage

AYIKPA KACOUTCHY JEAN : Enseignant - Chercheur



Table des matières

Introduction	3
I - 1- La table à adressage direct	4
II - Application 1:	6
III - 2- Tables de hachage	7
IV - Application 2:	g
Solutions des exercices	10

Introduction



De nombreuses applications font appel à des ensembles dynamiques qui ne supportent que les opérations de dictionnaire INSÉRER, RECHERCHER et SUPPRIMER.

Par exemple, un compilateur doit gérer une table de symboles, dans laquelle les clés des éléments sont des chaînes de caractères arbitraires qui correspondent aux identificateurs du langage.

Une table de hachage représente une autre façon de stocker des données. c'est est une structure de données permettant d'implémenter efficacement des dictionnaires. Bien que la recherche d'un élément dans une table de hachage puisse être aussi longue que la recherche d'un élément dans une liste chaînée.

En pratique le hachage est très efficace. Avec des hypothèses raisonnables, le temps moyen de recherche d'un élément dans une table de hachage est O(1).

1- La table à adressage direct



Définition : 1.1- Définition

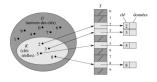
 $L'adressage \ direct$ est une technique simple qui fonctionne bien lorsque l'univers U des clés est raisonnablement petit. Supposons qu'une application ait besoin d'un

ensemble dynamique dans lequel chaque élément possède une clé prise dans l'univers $U=\{0,1,\ldots,m-1\}$, où m n'est pas trop grand. On supposera que deux éléments

ne peuvent pas avoir la même clé.

Pour représenter l'ensemble dynamique, on utilise un tableau, aussi appelé table à adressage direct, T [0 . . m-1], dans lequel *chaque position, ou alvéole, correspond à*

une clé de l'univers U



Implémentation d'un ensemble dynamique à l'aide de table à adressage direct T [0,..., 9].

Chaque clé appartient a $U = \{0, ..., 9\}$ et correspond à un indice de T.

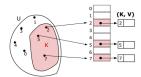
 $K = \{2, 3, 5, 8\}$ est l'ensemble des clés réelles et détermine les alvéoles de la table qui contiennent des pointeurs sur les éléments. Les autres alvéoles contiennent NIL.

Plus généralement, *l'alvéole k* pointe sur un élément de *l'ensemble ayant pour clé k. T[k]* \leftarrow *NIL* si l'ensemble ne contient aucun élément de clé k.

1.2- Algorithmes associés

- RECHERCHER-ADRESSAGE-DIRECT (T, k)
 renvoyer T[k]
- INSERER-ADRESSAGE-DIRECT (T, x) $T[clé[x]] \leftarrow x$
- SUPPRIMER-ADRESSAGE-DIRECT(T, x) $T[clé[x]] \leftarrow NIL$

Remarque: 1.3- Remarque



L'inconvénient de l'adressage direct est évident : si l'univers U est grand, gérer une table T de taille U peut se révéler compliqué, voire impossible, compte tenu de

la mémoire généralement disponible dans un ordinateur.

Par ailleurs, l'ensemble K des clés réellement conservées peut être tellement petit comparé à U que la majeure partie de l'espace alloué pour T est gaspillé.

Lorsque l'ensemble K des clés stockées dans un dictionnaire est beaucoup plus petit que l'univers U de toutes les clés possibles, une table de hachage requiert moins

de place de stockage qu'une table à adressage direct.

Application 1:



Exercice [solution n°1 p.10]

L'adressage direct fonctionne bien :

- O lorsque l'univers U des clés est raisonnablement grand
- O lorsque l'univers U des clés est petit
- O lorsque l'univers U des clés est grand

Exercice [solution n°2 p.10]

Un alvéole est:

- O une représentation d'une table
- O est l'ensemble dynamique d'un tableau
- O est une position de l'adressage direct

2- Tables de hachage



2.1- Généralité

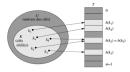
Une table de hachage est une structure de données qui permet une association clé-valeur, c'est-à-dire une implémentation du type abstrait tableau associatif; en particulier, l'implémentation d'une table des symboles lorsque les clés sont des chaînes de caractères.

Avec *l'adressage direct*, un élément de *clé* k est conservé dans *l'alvéole* k. Avec *le hachage*, cet élément est stocké dans *l'alvéole* h(k); autrement dit, on utilise une

fonction de hachage h pour calculer l'alvéole à partir de la clé k.

h établit une correspondance entre l'univers U des clés et les alvéoles d'une table de hachage.

On dit qu'un élément de clé k est haché dans l'alvéole h(k); on dit également que h(k) est la valeur de hachage de la clé k.



- Le but de *la fonction de hachage* (h) est de réduire l'intervalle des indices de tableau à gérer. Au lieu de U valeurs, il suffit de gérer m valeurs. Les besoins en stockage sont réduits en conséquence.
- L'inconvénient de cette idée est que *deux clés peuvent être hachées vers la même alvéole*, entraînant ainsi une collision. Heureusement, il existe des techniques efficaces pour résoudre les conflits créés par les collisions.

2.2-Résolution de collisions par chaînage

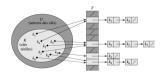
Avec le chaînage, on place dans une liste chaînée tous les éléments hachés vers la même alvéole. L'alvéole j contient un pointeur vers la tête de liste de tous les éléments hachés vers j ; si aucun n élément n'est présent, l'alvéole j contient NIL.

Les opérations de dictionnaire sur *une table de hachage T* sont faciles à implémenter lorsque les collisions sont résolues par chaînage.

INSERER-HACHAGE-CHAÎNEE(T, x): insère x en tête de la liste T[h(clé[x])]

RECHERCHER-HACHAGE-CHAÎNEE(T, k): recherche un élément de clé k dans la liste T[h(k)]

SUPPRIMER-HACHAGE-CHAÎNÉE(T, x): supprime x de la liste T[h(clé[x])]



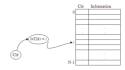
2.3- Fonctions de Hachage

Une fonction de hachage est une fonction qui transforme une donnée (la clé) en un entier (l'indice dans le tableau).

Formellement : h(k) = i, elle établit une correspondance entre un *couple(clé,valeur) et un indice du tableau*. Le *couple se trouvera* à T[i].

Principe:

C'est une technique très utilisée en informatique, elle se base sur une fonction h appelée fonction de hachage ou de hashcoding, qui appliquée à la clé fournit l'indice correspondant dans la table.



On trouve plusieurs types de fonctions utilisées en Hashcoding :

- Si la clé est une chaîne de caractères (nom)

h(clé) = CodeASCII (1er car) + Code ASCII(2eme `car) mod N

CodeASCI : ramène la valeur décimale correspondant à la table ASCII voir le lien : http://www.table-ascii.com/

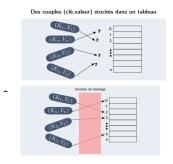
- Si clé est une valeur numérique.

 $h(clé) = clé \mod N$

La fonction de hachage doit donner des valeurs entières dans l'intervalle des indices de la tables : 0 <=h(cle') <= N.

En plus, cette fonction doit être la plus distribuée possible sur cet intervalle, pour que les informations ne se concentrent pas dans une partie de la table.

- Illustration:



Une bonne fonction de Hachage vérifie (approximativement) l'hypothèse de Hachage uniforme : chaque clé a autant de chance d'être hachée dans une quelconque des m alvéoles.

Application 2:



Exercice [solution n°3 p.10]

Une table de hachage

O est une structure de données permettant d'implémenter des valeurs à indice numérique

0

est une structure de données qui permet l'implémentation d'une table des symboles lorsque les clés sont des chaînes de caractères.

O est une structure de données permettant d'utiliser les listes

Exercice [solution n°4 p.10]

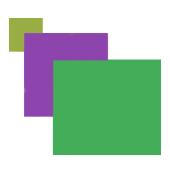
Une fonction de hachage est :

- O établit une correspondance entre un association clé, valeur et un indice du tableau
- O établit une correspondance entre un couple(clé,valeur) et une chaîne

établit une correspondance entre un couple(clé,valeur) et une liste

O chaînée

Solutions des exercices



> Solution n°1

L'adressage direct fonctionne bien :

- O lorsque l'univers U des clés est raisonnablement grand
- \odot lorsque l'univers U des clés est petit
- O lorsque l'univers *U* des clés est grand

> **Solution** n°2

Un alvéole est:

- O une représentation d'une table
- O est l'ensemble dynamique d'un tableau
- est une position de l'adressage direct

> **Solution** n°3

Une table de hachage

0

- O est une structure de données permettant d'implémenter des valeurs à indice numérique
- est une structure de données qui permet l'implémentation d'une table des symboles lorsque les clés sont des chaînes de caractères.

O est une structure de données permettant d'utiliser les listes

> Solution n°4

Une fonction de hachage est :



- établit une correspondance entre un association clé, valeur et un indice du tableau
- O établit une correspondance entre un couple(clé,valeur) et une chaîne établit une correspondance entre un couple(clé,valeur) et une liste
- O chaînée