VARI 8

LE HACHAGE

I TABLES DE HACHAGE

Utilisées pour stocker de grandes quantités d’informations

Recherche rapide d’un élément

Tableau [0,…,m-1]

Indice dans la table calculé en fonction de la clef de l’élément stocké à cet endroit

PRINCIPE DU HACHAGE

Fonction

$$h : \{\text{clés}\} \longrightarrow [0,..,m-1]$$

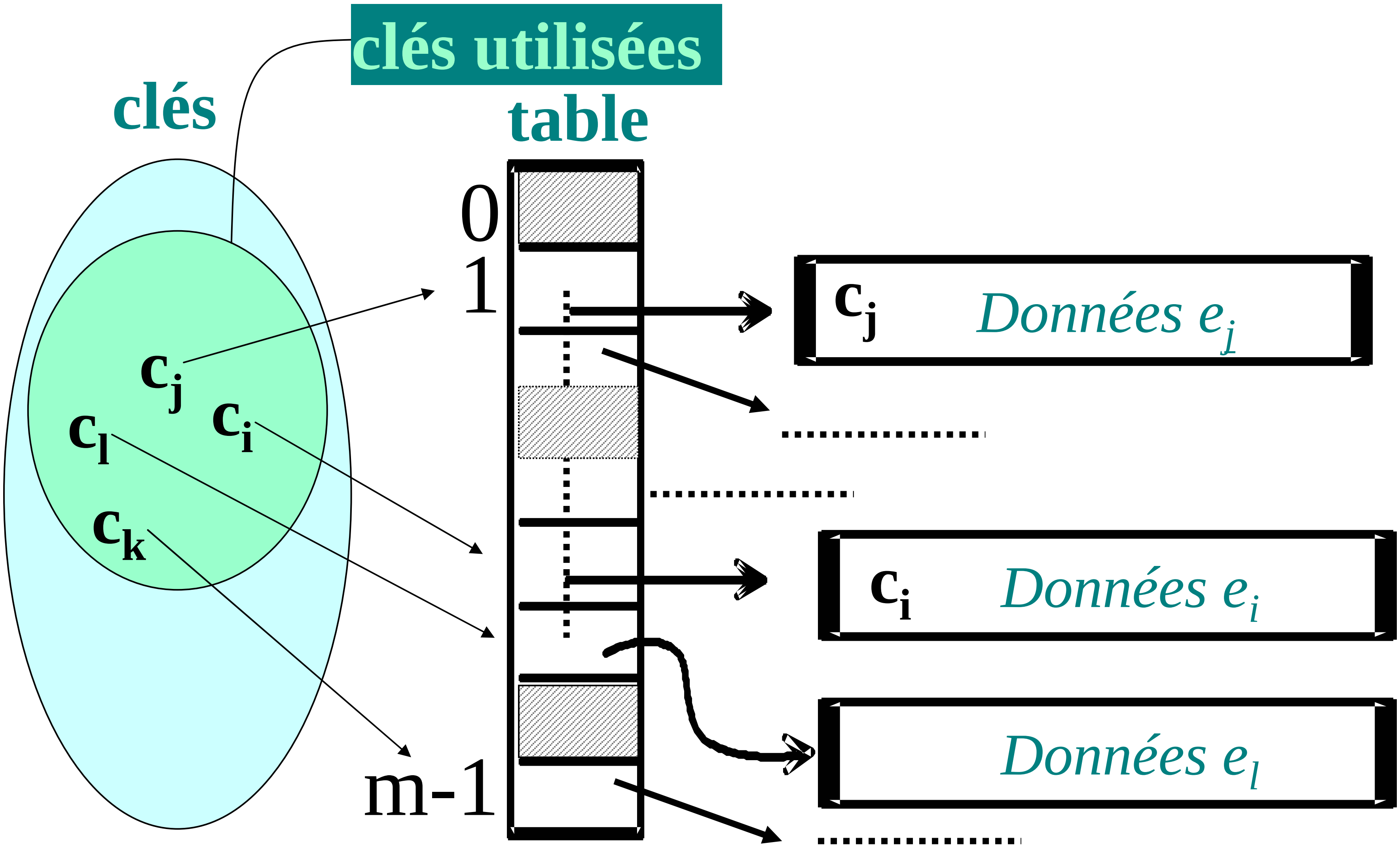
$$c \longrightarrow h(c) \qquad \textit{accès direct dans la table}$$

Problème car on peut avoir h non injective:

$$c \neq c' \quad \text{et} \quad h(c) = h(c')$$

Cas simple: adressage direct

Fonction h injective: une clé ⟶ une place



COLLISION

h non injective

$E = \{\text{éléments}\} = \{e_1, e_2, .., e_i, .., e_n\}$

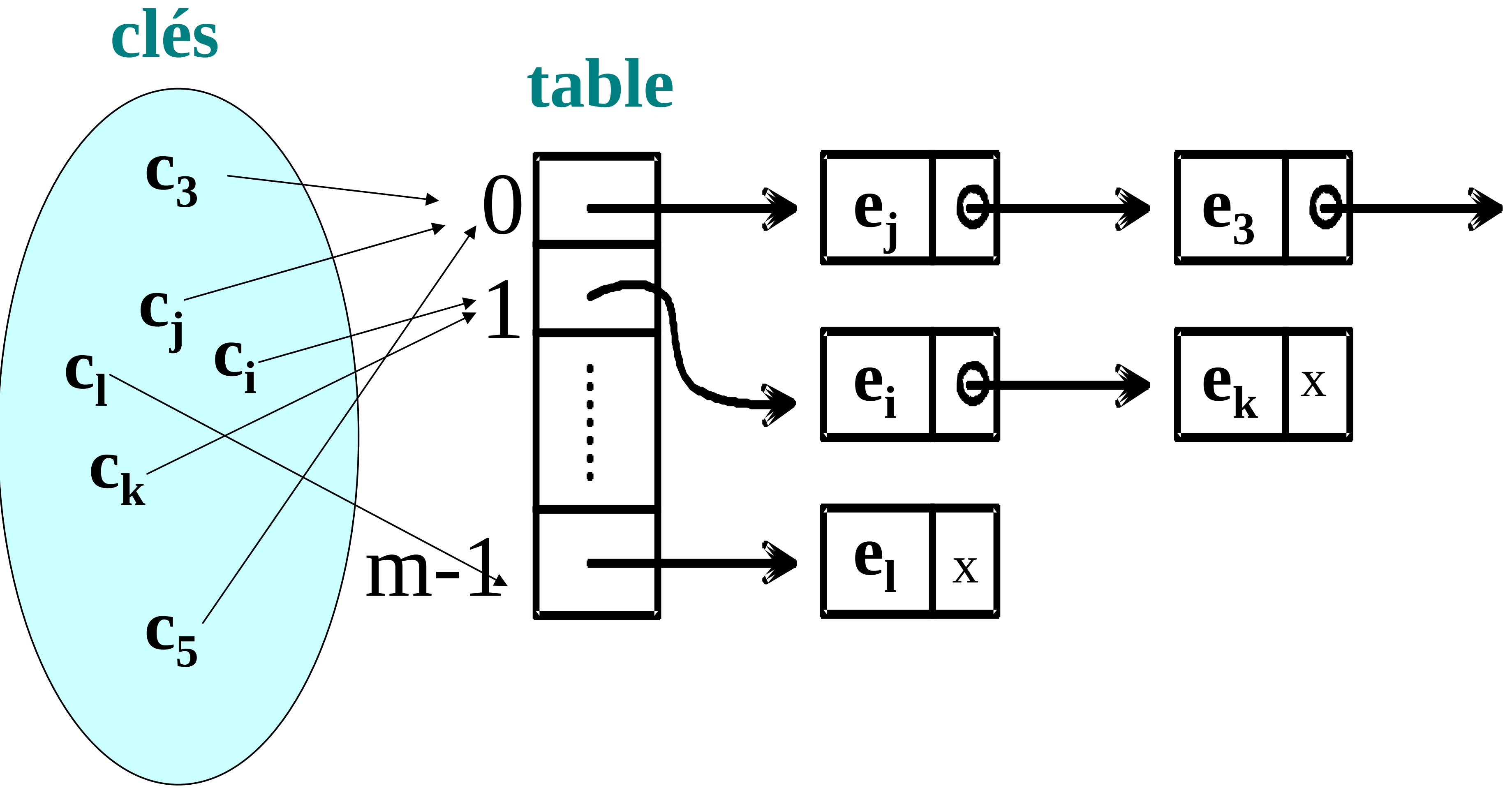
$$E \longrightarrow \{\text{clés}\} \longrightarrow [1,2,..,n]$$

$$e_i \longrightarrow \text{clé}(e_i) \longrightarrow h(\text{clé}(e_i))$$

Place de l'élément e_i : $T(h(\text{clé}(e_i)))$

Collision si $h(\text{clé}(e_i)) = h(\text{clé}(e_j))$

adressage indirect et chaînage



n clés

m places

- Temps de recherche de l'élément de clé c

=

Calcul de l'indice h(c) dans T: O(1) +

\+ Parcours de la liste chaînée : $O(L_{h(c)})$

- Idéal: éviter les collisions
- Réaliste: limiter les collisions

- Important: choisir une « bonne » fonction de hachage

II

FONCTIONS DE HACHAGE DE BASE

- **une fonction de hachage transforme**
  - **la clé d'un élément en**

  ⇨ **l'indice dans la table de hachage**
- **cette fonction doit être**
  - **déterministe**
  - **la plus uniforme possible**

    (occupation "régulière" de la table)
  - **facile à calculer**
- **fonction de hachage uniforme**:

$\forall$ **x** $\in$ **E et** $\forall$ **i** $\in$ **[1, m] Proba (h(x) = i) = 1/m**

Avis

- La suite des transparents est manuscrite

- Nous vous prions d'accepter nos excuses.

- EXTRACTION DE BITS

une clé $\longrightarrow$ suite de bits $\longrightarrow$ extraction de p bits $\longrightarrow$ nombre entre 0 et $2^p - 1$

ex: C_1 01100 $\boxed{01101}$ 10011 ... — $\longrightarrow T(13)$

$2^4 \; 2^3 \; 2^2 \; 2 \; 1$

$\downarrow$

13

C_2 1010 $\boxed{01101}$ 01011 ....

COLLISION entre C_1 et C_2

- COMPRESSION DE BITS

METHODE DE LA DIVISION

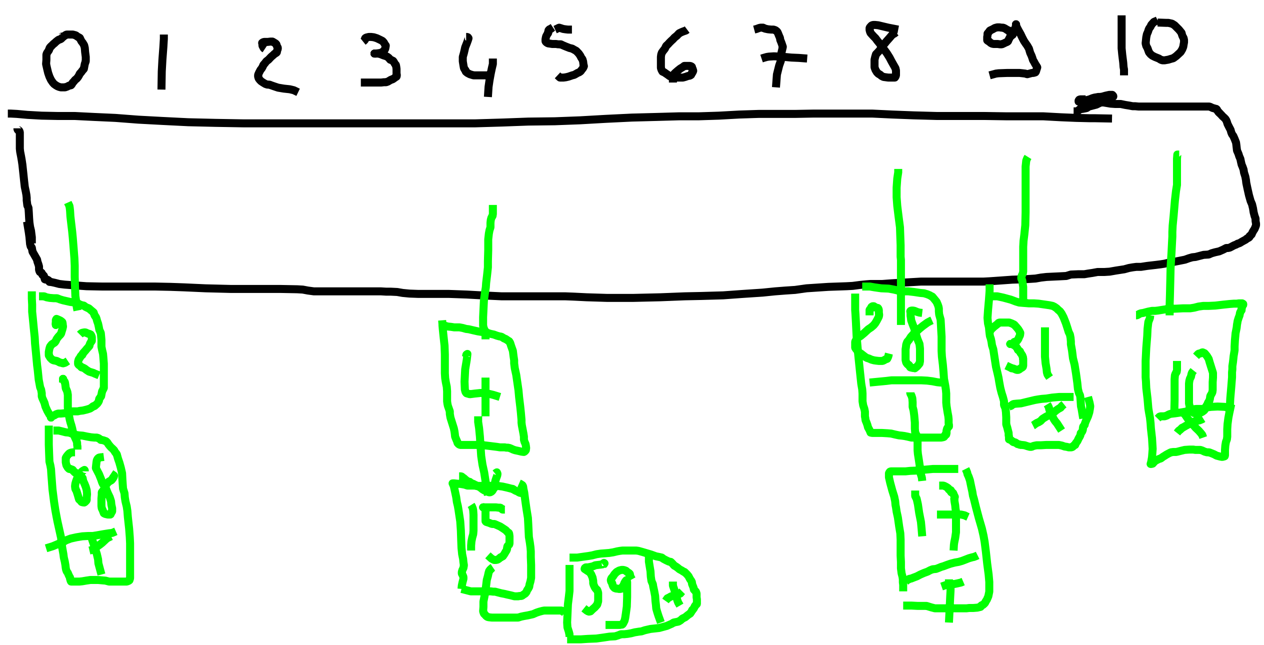
$C \to \mathbb{N}$

$$h(c) = c \bmod m$$

c	10	22	31	4	15	28	17	88	59
	10	0	9	4	4	6	6	0	4

$m = 11$

T



METHODE DE LA MULTIPLICATION

$$\underbrace{C \times \underset{r}{\text{constante}}}_{}$$ $0 < r < 1$

$C.r \rightarrow$ partie fractionnaire $C.r \bmod 1$

$= C.r - \lfloor C.r \rfloor_{\text{partie entière}}$

$h(c) = \lfloor (c.r \bmod 1) \times m \rfloor$

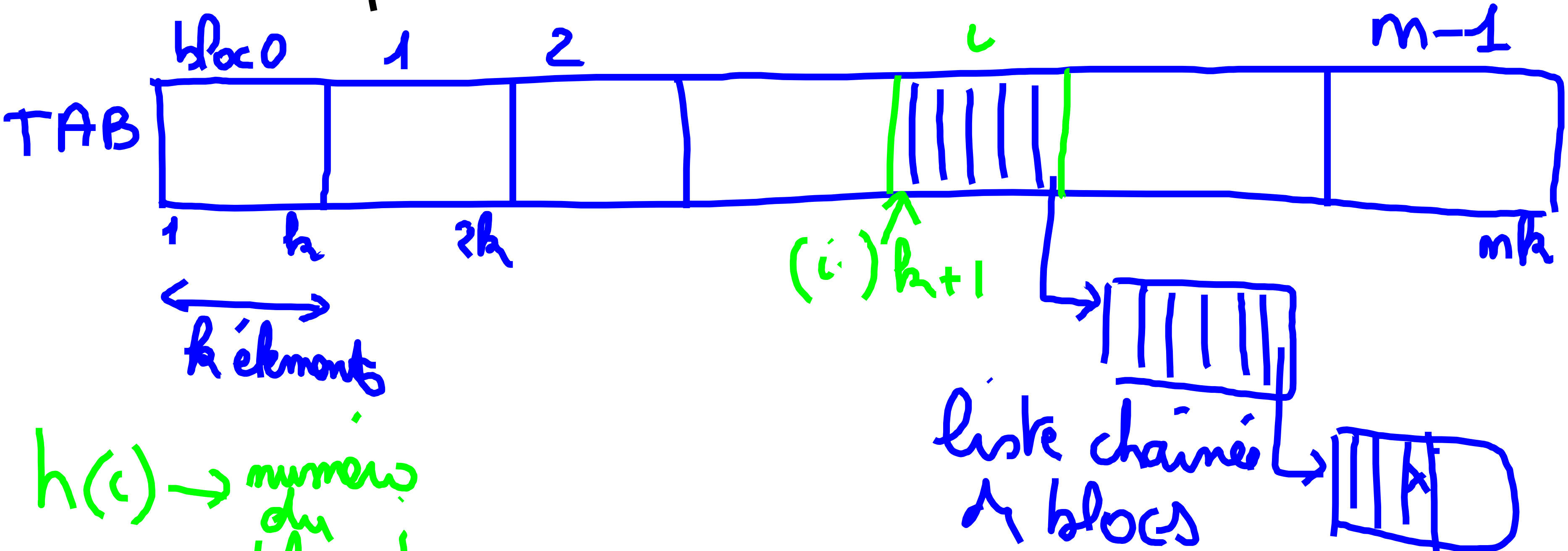
Knuth : $r \approx \frac{\sqrt{5}-1}{2} \simeq 0,6180339887$

$c = 123456 \quad m = 10000$

$h(c) = \lfloor 0,0041151 * 10000 \rfloor = \lfloor 41,151 \rfloor = 41$

III IMPLANTATION

bloc disque $\longleftrightarrow$ une liste chaînée k éléments

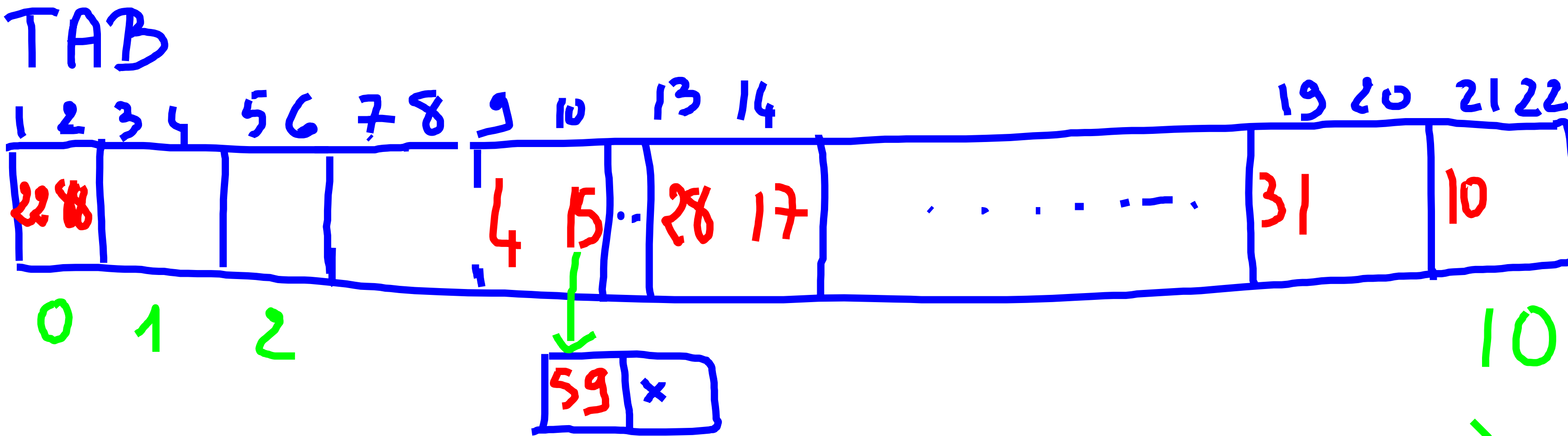


$h(c) \rightarrow$ numero du bloc i

$\rightarrow$ indice dans TAB $(i)k+1$

Exemple

Blocs de "2" éléments



11 blocs de taille 2

$h(c) = c \bmod m$

$TAB(2 * i + 1)$

Recherche de l'élément de clé c dans la table de hachage

- Calculer $h(c)$ $\quad$ .. $h(c) = c \bmod m$
- Calculer ind_bloc : $I = ((h(c)-1) \times k) + 1$
- Rechercher dans le premier bloc

Tant que $I \leq h(c) \times k$ faire
    si TAB[I].clé = c alors TROUVÉ
    sinon si TAB[I].dernier = vrai alors ABSENT
        sinon I := I+1
    fin si
fait

- Recherche dans le bloc suivant si il existe

Tant que ∃ bloc suivant faire
    Ca adresse bloc suivant

- Tant que ∃ bloc suivant faire

    lire adresse bloc ;   I := 1 ;

    Tant que I ≤ k faire

        si Bloc[I].clé = c alors TROUVÉ

        sinon si Bloc[I].dernier = vrai alors ABSENT

            sinon I := I+1 ;

            fin ;

        finsi

    fait

fait

- ÉLÉMENT ABSENT

    FIN

IV ADRESSAGE OUVERT

IV.1 PRINCIPE

Pas de chaînage

1 case → nil (rien)
ou
→ un élément

Idée : $c \to h(c) =$ 1ère place si libre OK
sinon calcul d'une 2e place
si libre OK
sinon 3e place
etc.

SONDAGE

Règle précise

Complet (1 fois par chaque place)

exemple suite de places par clé

$c_1 \rightarrow$	1	3	5	0	2	4
$c_2 \rightarrow$	3	5	2	1	0	4
$c_3 \rightarrow$	3	0	1	2	4	5
$c_4 \rightarrow$	1	3	5	0	4	2
i	1	2	3	4	5	6

ième place testée

PAS DE SUPPRESSION

TAB	0	1	2	3	4	5
	c_3	c_1		c_2		c_4

Recherche dans le même ordre

Problème pour recherche si suppression

AVEC SUPPRESSION

TAB

0	1	2	3	4	5
C_3	C_1	nil	~~C_2~~ ôté	nil	C_4

Retrait d'un élément → "ôté"

C_2 retiré

Recherche de C_4 :
1 non car C_1
3 non car ôté
5 oui

Insertion possible (en 3) si "ôté"

IV.2 Fonctions de hachage pour adressage ouvert

$$h : \mathcal{C} \times \{0,1,\dots,m-1\} \longrightarrow \{0,1,\dots,m-1\}$$

clé numéro de l'essai

$c \rightarrow h(c,0)$ puis $h(c,1)$ puis $h(c,2)$ etc.

il faut que

$\langle h(c,0); h(c,1); \dots ; h(c,m-1) \rangle$ permutation de $\{0,1,2,\dots,m-1\}$

a) SONDAGE LINEAIRE

$h_0 : \mathcal{C} \to \{0, 1, \ldots, m-1\}$ par ex $h_0(c) = c \bmod m$

$h(c,i) = (h_0(c) + i) \bmod m$ $\quad i = 0, 1, \ldots, m-1$

$h(c, i+1) = (h_0(c) + i + 1) \bmod m$

$= [(h_0(c,i) + 1] \bmod m$

On sonde les places adjacentes :

$T(h_0(c))$ puis $T(h_0(c)+1)$ ......, $T(m-1)$, $T(0)$, $T(1)$

jusqu'à $T(h_0(c) - 1)$

Exemple

0	1	2	3	4	5	6	7	8	9	10
22	88			4	15	28	17	59	31	10

Phénomène de "grappes fortes"

→ autre solution

hachage quadratique

$$h(c,i) = \left[h_0(c) + a_1 i + a_2 i^2\right] \bmod m$$

a_1 et a_2 constantes à choisir de façon à avoir la permutation

$$h(c,i) = \left[\underbrace{h_0(c) \bmod m}_{\text{calculé au début}} + \underbrace{(a_1 i + a_2 i^2) \bmod m}_{\text{calculé au début}}\right] \bmod m$$

Exemple

$a_1 = 1 \quad a_2 = 3$

i	$(a_1 i + a_2 i^2) \bmod 11$
1	$1 + 3 = 4$
2	3
3	8
4	2

0	1	2	3	4	5	6	7	8	9	10
22		88	17	4		28	59		53	10

pour 17 on sonde 6 puis $6+4$ puis $6+3$ puis

$6+8 = 14 \bmod 11 \rightarrow 3$ ok

grappes faibles $\quad$ $\hat{m}$ clé = $\hat{m}$ sondage

DOUBLE HACHAGE

$$h(c,i) = \left(h_0(c) + i\, h_1(c)\right) \bmod m$$

h_1 et h_2 = 2 fonctions de hachage

$h_1(c) = c \bmod m$ $\quad$ $h_2(c) = 1 + c \bmod (m-1)$

$$h_0(c) + i\, h_1(c) = c \bmod m + i + i\, c \bmod (m-1)$$

$$= c \bmod m + (i-1)\left(1 + c \bmod (m-1)\right) + 1 + c \bmod (m-1)$$

$$= h(c, i-1) + h_1(c)$$

exemple

$h_1(c) = c \bmod 11$

$h_2(c) = 1 + c \bmod 10$

	10	22	31	4	15	28	17	88	59
h_2	1	3	2	5	6	9	8	9	10
$h_0 = h_1$	10	0	9	4	4	6	6	0	4

0	1	2	3	4	5	6	7	8	9	10
22		59	17	4	15	28	88		31	10

15 → sondage 4 → 4+6 occ → $(10+6) \bmod 11$ → 5 ok

88 → 0 → 9 → $9+9 = 18 \bmod 11$ → 7 ok

Récapitulatif