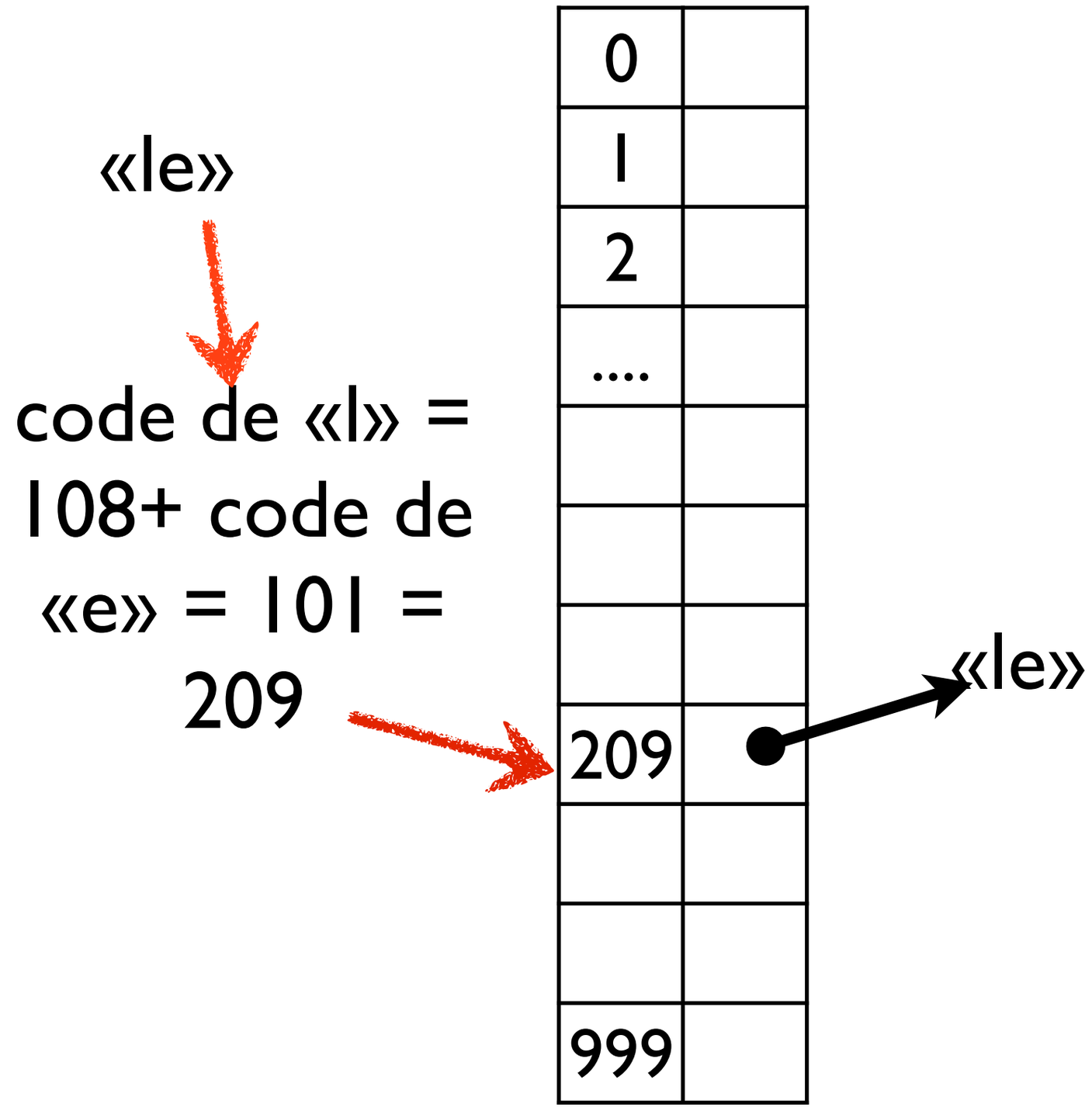
HashSet et HashMap: Tableaux de hachage

Principes

- On veut stocker et récupérer rapidement des données dans un tableau (par exemple des Strings)
- idée : on calcule une valeur entière à partir des données. Cette valeur donne la case du tableau où les données sont supposées se trouver
- cette valeur entière s'appelle fonction de hachage

Exemple simplifié

- On veut implanter un ensemble de Strings dans un tableau de 1000 cases
- fonction de hachage possible : somme des codes unicodes des caractères de la String

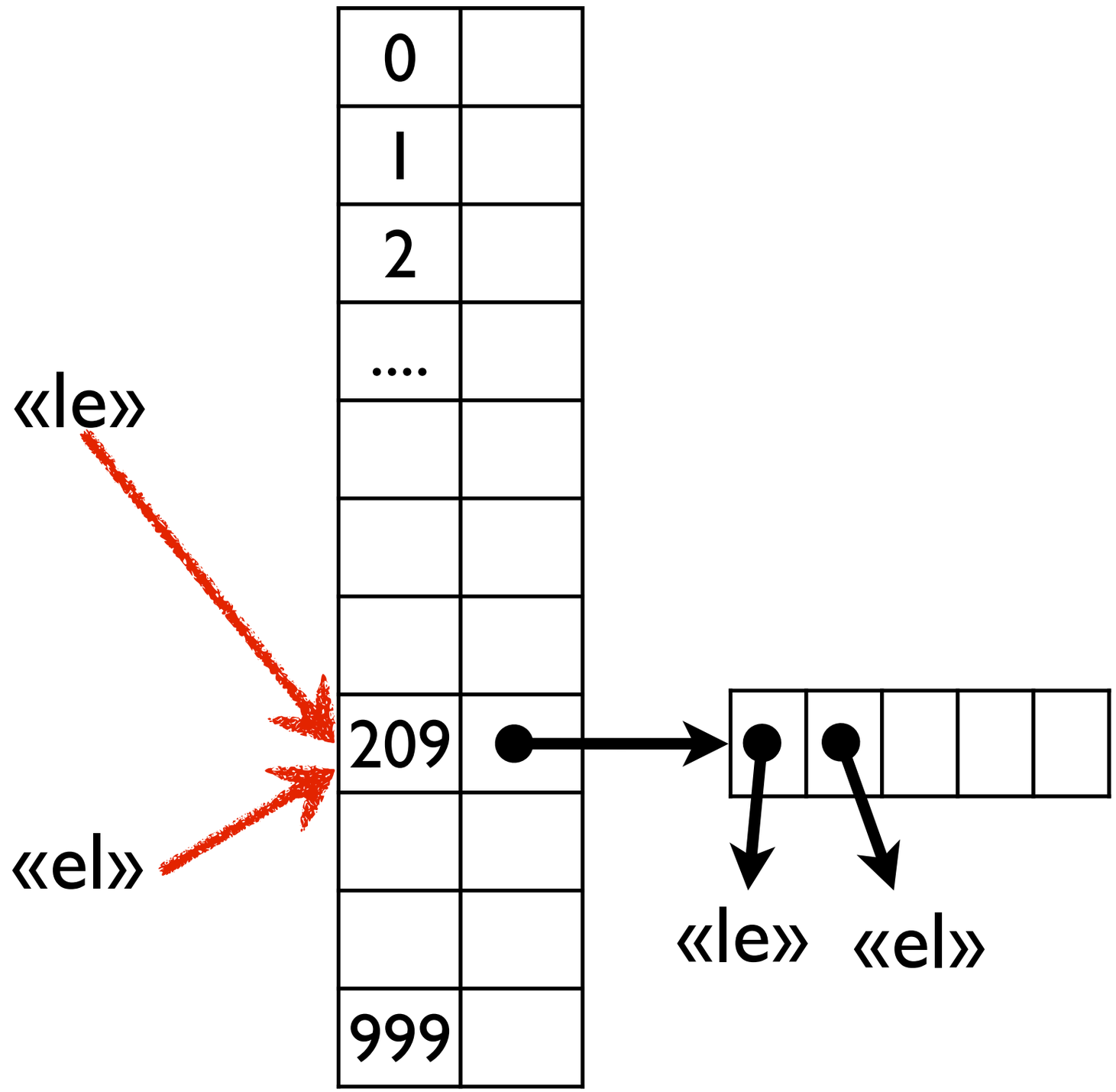


Problèmes

- Que se passe-t-il si la fonction de hachage est plus grande que la taille du tableau :
  - on prend le modulo
- Que se passe-t-il si (quand !) deux chaînes ont le même code de hachage («le» et «el» dans notre exemple ?)
  - collision. Plusieurs systèmes possibles...

Collision

Solution la plus simple : les cases du tableau contiennent des listes de valeurs et non pas les valeurs directement



Algorithme de recherche

- soit s à chercher dans une table de hachage t
- on calcule *h* le <<hashcode>> de s
- on prend sa valeur modulo t.length
- ça nous donne une case *i* où s est supposé se trouver
- dans la case i: parcourir la liste en comparant avec equals, pour voir si on trouve s

Algorithme d'insertion

- soit s à ajouter dans une table de hachage t
- on calcule *h* le hashcode de s
- on prend sa valeur modulo t.length
- ça nous donne une case *i* où s est supposé se trouver : i = s.hashCode() % t.length
- on parcourt la liste *l* qui est dans la case *i*, en comparant avec equals, pour voir si on trouve s
- si s n'est pas trouvée, on l'ajoute à *l*.

Conclusions Partielles

- pour 100000 valeurs dans un tableau de taille 1000 : en moyenne 100 valeurs par case
- intéressant si la fonction de hachage répartit uniformément les valeurs
- la classe Object fournit les méthodes hashCode() et equals() : on peut utiliser des HashSet avec toutes les classes java !
- Par défaut (dans Object) : hashCode dépend de *l'adresse* de l'objet. Les classes peuvent redéfinir la méthode si nécessaire.