

Hash code – Funcții de dispersie

Funcții de dispersie

- ▶ Obiect *hash-uibil* = care are cod **hash** asociat, returnat de metoda `__hash__()` (= cod de dispersie, valoare hash)

= un număr întreg cu proprietățile

Funcții de dispersie

- ▶ Obiect *hash-uibil* = care are cod **hash** asociat, returnat de metoda `__hash__()` (= cod de dispersie, valoare hash)

= un număr întreg cu proprietățile

- nu se poate schimba dacă obiectul nu se modifică (de obicei imutabil)
- două obiecte **egale au același hash code**
- este recomandabil ca două obiecte diferite să aibă valori hash diferite (compatibil cu `__eq__()`)

Funcții de dispersie

- ▶ Obiect *hash-uibil* = care are cod **hash** asociat, returnat de metoda `__hash__()` (= cod de dispersie, valoare hash)

= un număr întreg cu proprietățile

- nu se poate schimba dacă obiectul nu se modifică (de obicei imutabil)
- două obiecte **egale au același hash code**
- este recomandabil ca două obiecte diferite să aibă valori hash diferite (compatibil cu `__eq__()`)

Funcții de dispersie

Obiect



Număr întreg

dimensiune variabilă

plajă limitată de valori



apar coliziuni

Funcții de dispersie

```
t = (1,2)
```

```
print(hash(t)) #print(t.__hash__())
```

```
t = (1,[2,3])
```

```
print(hash(t)) # print(t.__hash__())
```

```
#TypeError: unhashable type: 'list'
```



Funcții de dispersie

- ▶ Funcția hash (de dispersie) trebuie aleasă astfel încât să se **minimizeze** numărul coliziunilor (obiecte diferite care produc aceleași hash-uri).
- ▶ Obiectele hash-uibile pot fi folosite ca chei pentru structuri de date indexate după chei (se va folosi pentru indexare codul lor hash)

Tabele de dispersie

- ▶ structură de date pentru căutare eficientă după chei hash-uibile
- ▶ căutare – în **medie** timpul $O(1)$ (defavorabil $O(n)$)
- ▶ un tablou

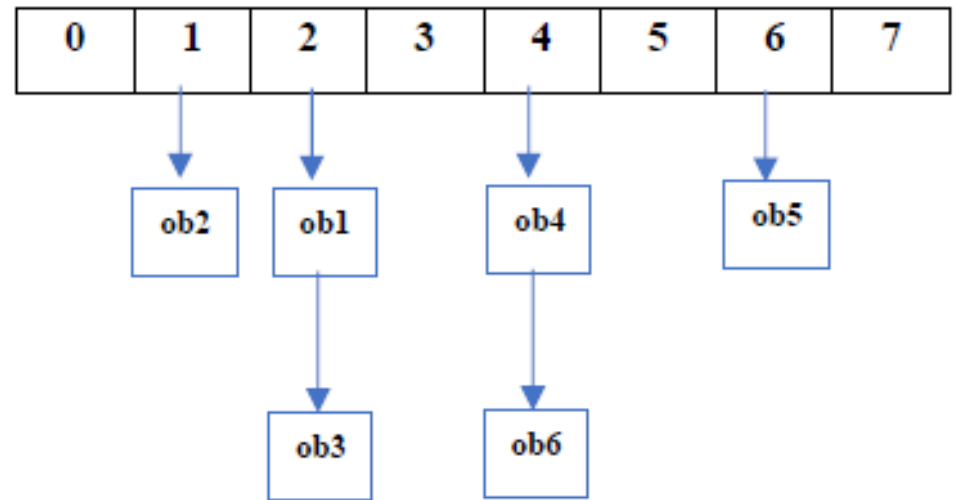
Tabele de dispersie

- ▶ un tablou
- ▶ Fiecărei poziții c din tabel îi corespunde un **bucket** (o listă) cu obiectele având indexul asociat codului hash (modulo numărul de bucketuri) egal cu c ;
- ▶ pot fi mai multe obiecte cu același index.

Tabele de dispersie

3 biți pentru memorarea codului hash

Obiect	Hash code (modulo numărul de bucketuri)
ob1 = "a"	2
ob2 = "bc"	1
ob3 = "ad"	2
ob4 = "m"	4
ob5 = "casa"	6
ob6 = "aab"	4



Tabele de dispersie

- ▶ Căutarea unui obiect în tabel se face, în mare, astfel:
 - se determină $c = \text{indexul asociat codului hash al obiectului (modulo numărul de bucketuri)}$
 - se accesează bucketul c
 - se caută în acest bucket obiectul (folosind pentru testare metoda `__eq__()`).

Este bine ca un bucket să fie mic

\leq puține coliziuni \leq algoritmi funcție de dispersie