

Problema 3.1: El hash tancat

Josep M. Ribó

22 de desembre de 2010

1 Objectius

- Discutir la forma d'implementació d'una taula (map) mitjançant la tècnica de la dispersió (o hash)
- Experimentar amb costos, apinyament, famílies de dispersió...

Per fer aquest problema, cal que us llegiu:

- Els apunts (apartats 3.2 i 3.4)

2 Funcions de hash i dimensió de la taula

Accediu a:

<http://griho2.udl.cat/josepma/eines/hash/principa.html>

Nota: Aquesta aplicació va ser dissenyada sobre una versió de Java més antiga. És possible que us trobeu amb alguns problemes de visualització. Aquests problemes no us han d'impedir treballar amb ella.

Per poder-la executar necessitareu el plugin de java al navegador

1. Pitgeu **Redimensionar ara**. Trieu un valor per B (1..501). En realitat, aquest valor no l'usarem per a res, però l'aplicació requereix que estigui inicialitzat abans que cap altra cosa.
2. Marqueu **Executar exemples**
3. Seccioneu **Tipus hash: tancat**
4. Seccioneu **Funció hash: Mòdul1**

La funció de hash **Mòdul1** és la següent:

$$h(k) = \left(\sum_{i=0}^{i<|k|} \text{codi}(k_i) \right) \bmod B$$

Aquesta funció pateix un problema a l'hora de dispersar les claus. En particular, totes les claus formades per permutacions dels mateixos caràcters aniran a parar al mateix índex (exemple: "anton" i "notan").

5. Seccioneu **Família funcions redispersió: Lineal**

6. Pitgeu Inicialitzar

Seleccioneu un factor de càrrega (α)=0.8 i $n= 460$ claus a introduir

En algunes resolucions el valor de n costa una mica d'introduir perquè el camp de text on s'ha d'introduir apareix tapat per un element del quadre de diàleg. Amb tot i això és possible introduir-lo. **Pitgeu la tecla DEL del teclat varies vegades** per tal d'assegurar que el cursor queda a l'inici del camp de text on heu d'escriure el valor de n i després escriviu aquest valor n

7. Observeu la gràfica que apareix (recordeu que α = factor de càrrega)

Aquesta gràfica indica el nombre de proves que ha calgut fer per col·locar cadascuna de les claus a mesura que s'ha anat incrementant el factor de càrrega fins arribar a aproximadament 0.8.

8. Pitgeu Reiniciar-ho tot

Repetiu la prova amb les configuracions següents:

- Hash tancat, funció mòdul 1, redispersió quadràtica, $\alpha=0.8$, $n= 460$
- Hash tancat, funció mòdul 2, redispersió lineal, $\alpha=0.98$, $n= 460$

Nota: La funció de hash *mòdul 2* és la presentada a classe:

$$h(k) = \left(\sum_{i=0}^{i < |k|} \text{codi}(k_i) * b^i \right) \bmod B$$

- Hash tancat, funció mòdul 2, redispersió quadràtica, $\alpha=0.98$, $n= 460$
- Hash tancat, funció mòdul 2, redispersió doble, $\alpha=0.98$, $n= 460$
- Repeteix les proves amb $\alpha=0.7$

1. Per cada situació comenta el comportament de l'eficiència (nombre d'assajos o proves) a mesura que α creix.
2. Quina et sembla la millor funció de hash de les provades?
3. Quina família de funcions de redispersió va millor en aquests exemples?
4. Quin factor de càrrega màxim et sembla acceptable (sense que hi hagi una gran degeneració del rendiment)?
5. En quin cas es pot apreciar d'una manera més clara el fenomen de l'apinyament primari (les "butifarres")?