

CS 351 DATA ORGANIZATION AND MANAGEMENT

Homework 3 (Updated)

Date Given : November 20, 2011

Date Due : November 24, 2011

Important Notes: 1. Please submit the Homework to Room EA 511 on the due date by 5:00 pm – on the same day you can also give it to me in the classroom- (no late submission will be accepted). 2. Answer the question in the order they are given using a standard size paper. 3. Handwritten submissions are accepted; however a word document is preferred and appreciated. 4. Staple all papers and write your name and section on them. 5. When needed assume/use the parameters/values of IBM 3380.

1. Using a bucket factor of two ($Bkfr = 2$, no. of records per bucket is 2) and the hash function $HF(key) = \text{mod}(key, 3)$ and bucket chaining enter into an empty static hash table (file) the following values: (27, 5, 18, 30, 10, 32, 38).
 - a. Draw a simple figure and show the file contents. Use separate chaining (see Fig. 6.1 in Salzberg).
 - b. How many disk accesses are required in order to retrieve the record 30? How much time is needed? (Assume that bucket size is the same as block size.)
 - c. How many disk accesses are required in order to retrieve the record 32? How much time is needed? (Assume that bucket size is the same as block size.)
 - d. What is the average number of accesses for an unsuccessful search?
2. Consider a Linear Hashing¹ environment with $Bkfr = 3$ and desired Lf (desired load factor) = $2/3$.

Update the file structure when the “current load factor of the file becomes $2/3$ and after adding 2 records.” (Note that this is the approach used in Salzberg’s book, so make sure that you follow that approach that we also considered in the class.)

Start with $h = 1$, $bv = 0$, and use $\text{mod}(key, 17)$ to obtain the pseudo key values.

Show the file growth process by inserting records in the following order (each number represents a record): 10, 20, 3, 7, 13, 14, 18, 21, 25, 16, 38, 30, 37, 35, ~~25~~, 23. (~~delete 25~~)

After each insertion please clearly indicate the values of h and bv (boundary value, another name for it is split pointer: sp), and the current load factor (Lf) of the file.

Comment: Note that h is the hashing level and as the file space grows higher values of h are used. In the Salzberg book the author uses the symbol k instead of h .

3. In a Linear Hashing file environment $bv = 3$ and $h = 4$. Give the values for the following items.
 - a. Number of blocks hashed at level h .
 - b. Number of blocks hashed at level $h+1$.
 - c. The binary address of the last bucket of the file.
 - d. The binary address of the last bucket of the file hashed at level 4.
 - e. Which bucket (give it in binary) do we need to access to find the record with the key = 54?
Use $\text{mod}(key, 37)$ to obtain the pseudo key and convert it to binary.
Note that if $(\text{remainder} < bv)$ we use $(h+1)$ no. of bits to determine the bucket no.
4. In a Linear Hashing file environment $bv = 200$ and $h = 10$. Give the values for the following items.
 - a. Number of blocks hashed at level h .
 - b. Number of blocks hashed at level $h+1$.
 - c. The binary address of the last bucket of the file.
 - d. The binary address of the last bucket of the file hashed at level ~~4~~ 10. (~~replace 4 with 10~~)
 - e. Which bucket (give it in binary) do we need to access to find the record with the key value 5178? Use $\text{mod}(key, 2039)$ to obtain the pseudo key and convert it to binary.

¹ Some linear hashing problem examples from our course web site:

1. http://www.cs.bilkent.edu.tr/~canf/CS351Fall2011/Fall10Quizzes/cs351_f10_q3.pdf

2. <http://www.cs.bilkent.edu.tr/~canf/CS351Fall2011/Fall09HWs/CS%20351%20Hw3%20Sol2.pdf>