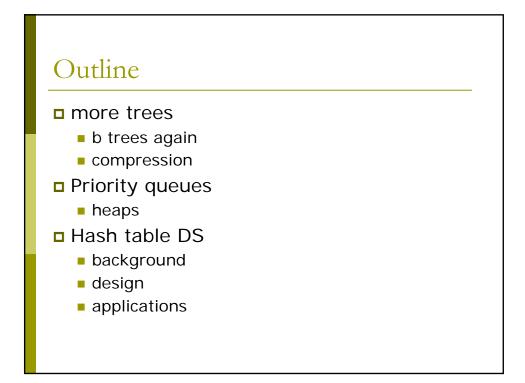
3137 Data Structures and Algorithms in C++

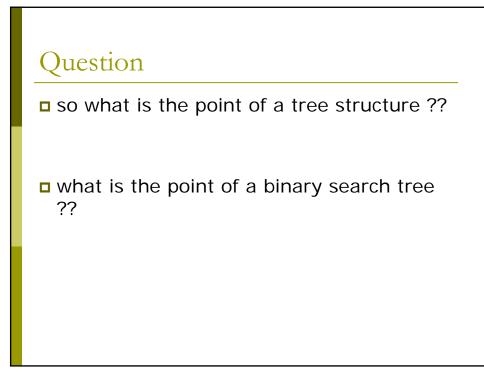
Lecture 5 July 19 2006 Shlomo Hershkop

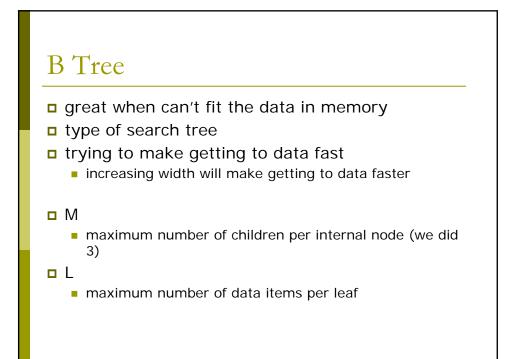
Announcements

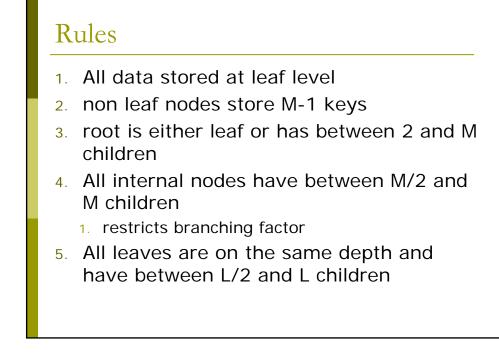
midpoint of semester today

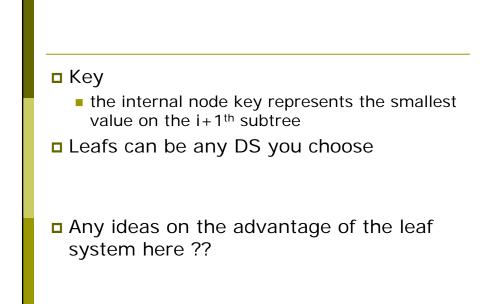
make sure you are ok with the material covered so far

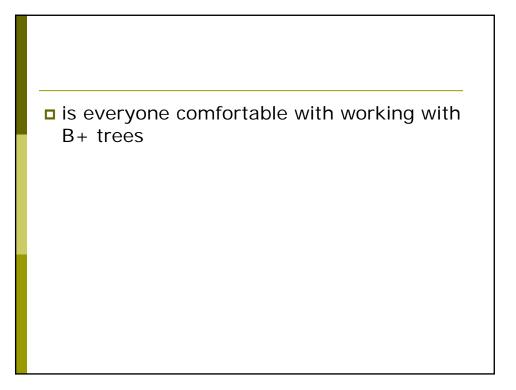


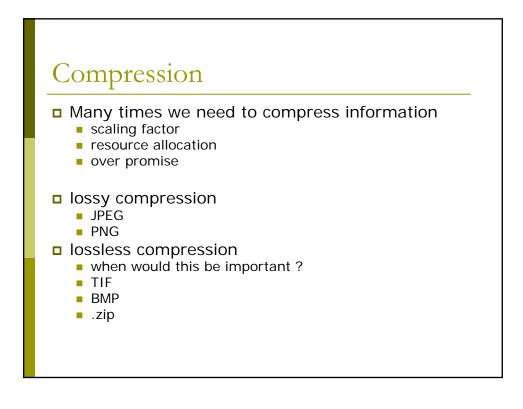


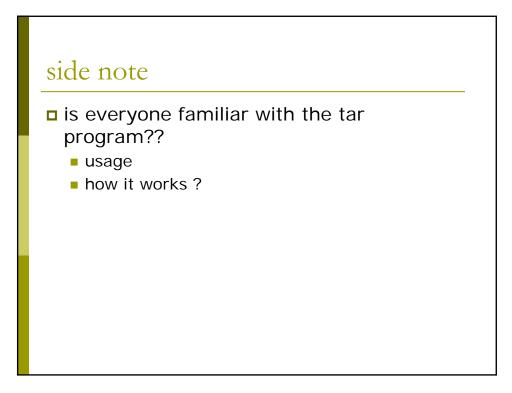


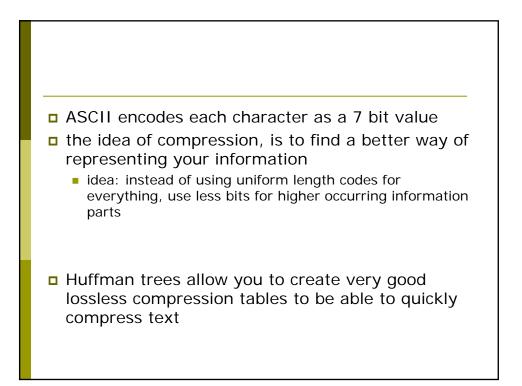


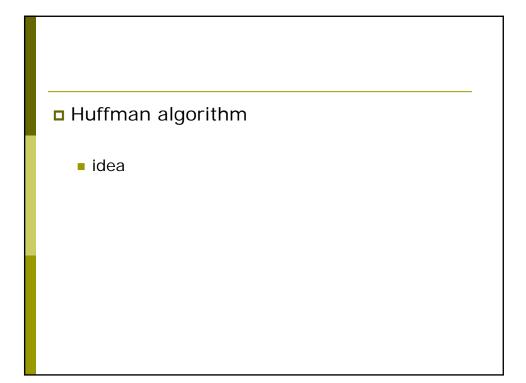














- 1. Create a frequency count of each of your characters in your file
- Start to build a binary tree always combining 2 lowest frequencies into one tree the resulting frequency is the combined frequencies
- 3. Going left is 0, going right is 1

Example

If I counted:
E = 29
A = 14
T = 10
B = 4
D = 2
C = 1

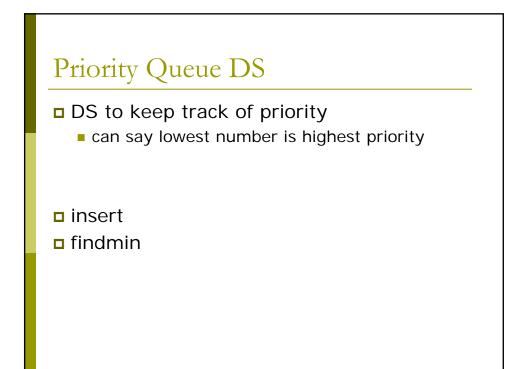
decompression

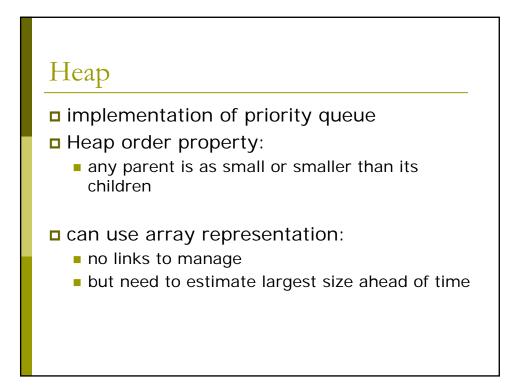
- So seeing a code, we simply run down the tree
- As soon as we hit a leaf, translate to that character

Compressing text

How would you use Huffman to compress text??

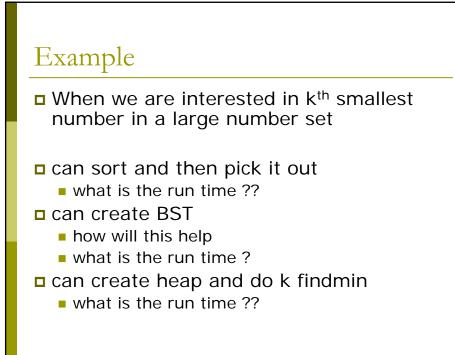
what about decompressing

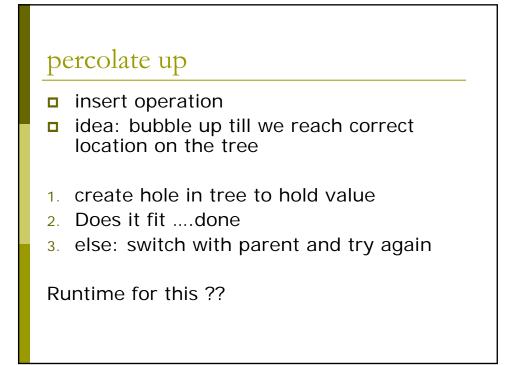




used everywhere

- service priority
- Operating systems juggling threads, processes and processors





findmin

■ so now we need to find min....

• what is the run time for finding the min ??

what if we want to find and delete ??

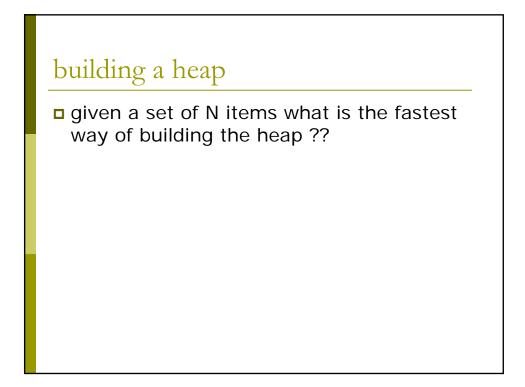
percolate down

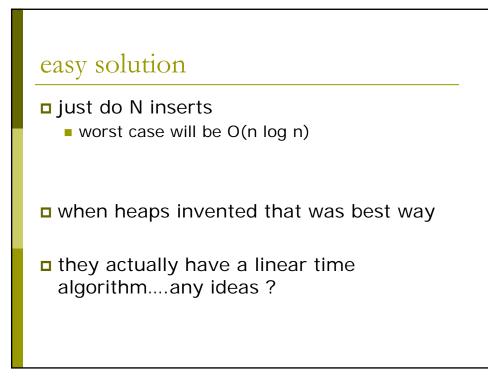
- pull out the root
- put hole
- swap with smaller child
- bubble down the hole

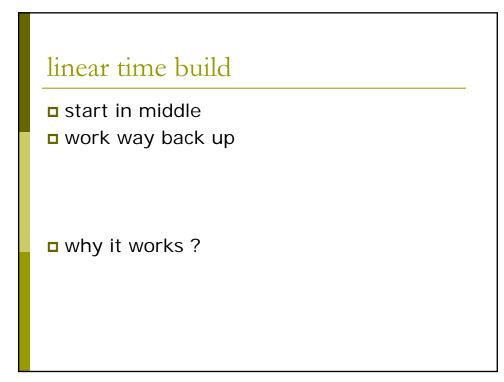
run time

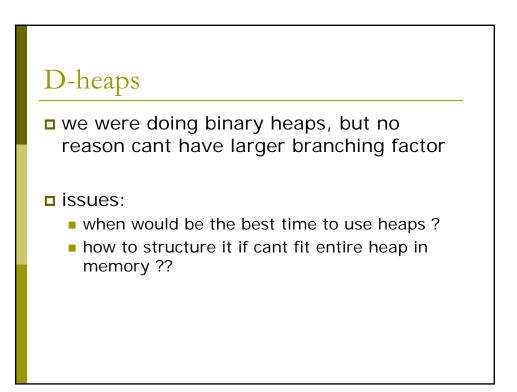
how long would it take me to find an item of specific priority ??

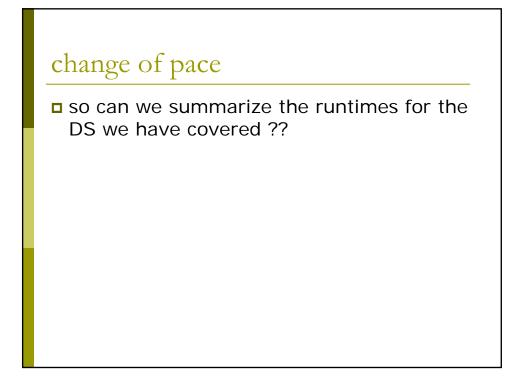
any ideas on how to help this ?

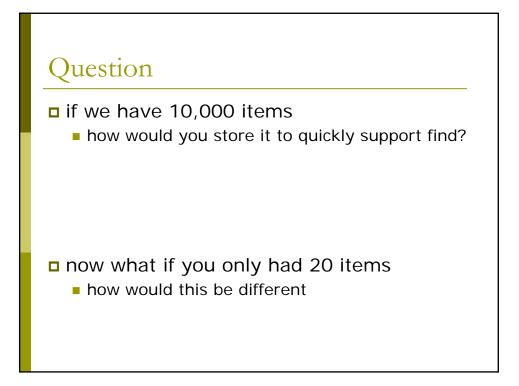


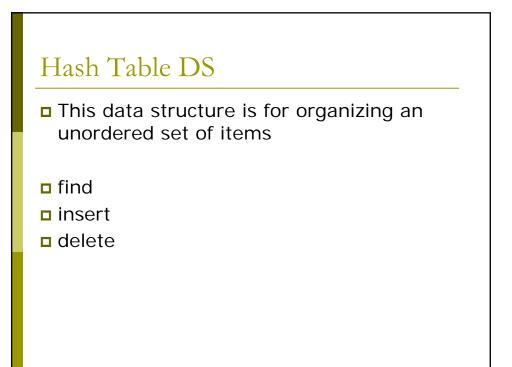


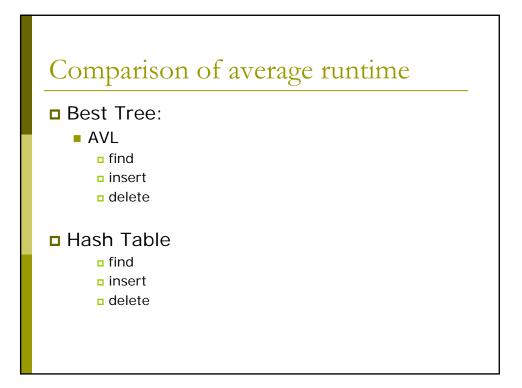


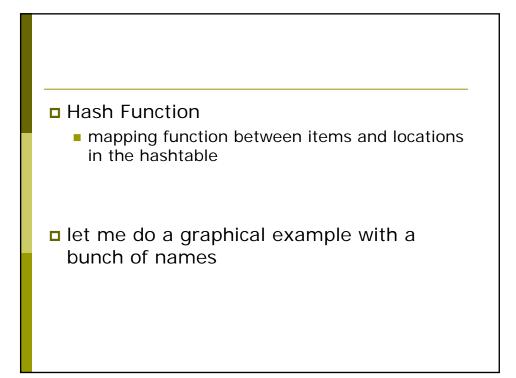








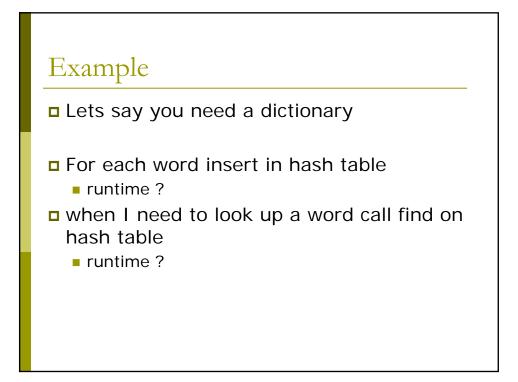


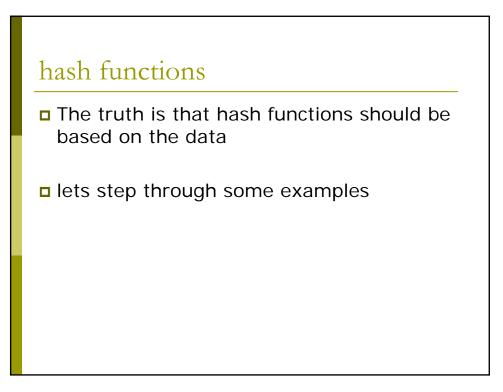


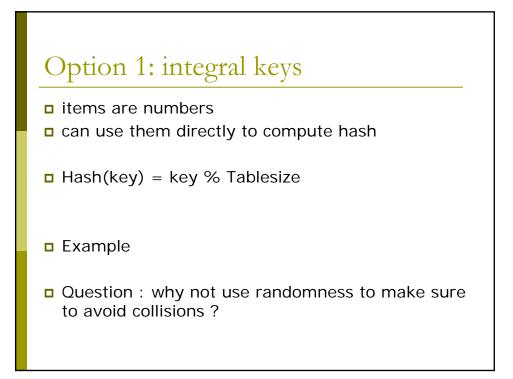
Issues

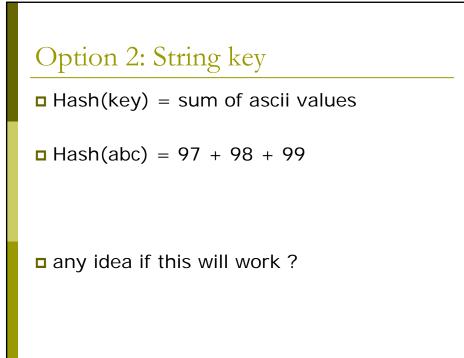
What hash function to use ?

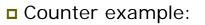
What do you do about collisions??









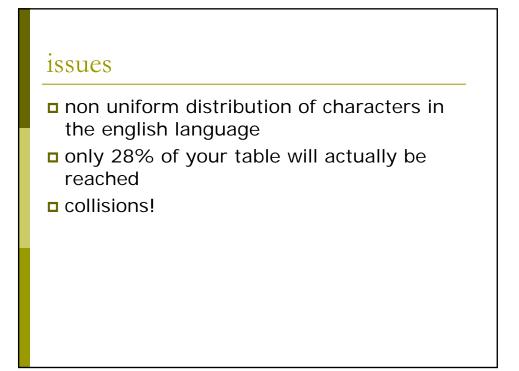


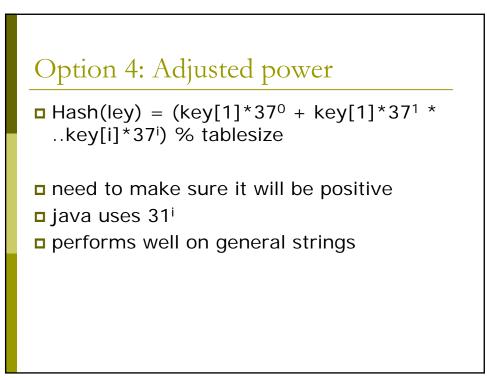
- dictionary
- tablesize 40,000
- what is the maximum word size
- what would be the max value returned by the hash ??

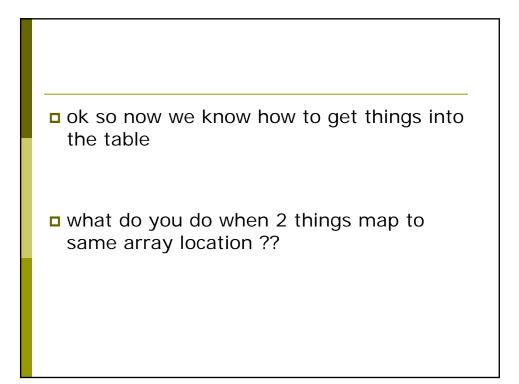


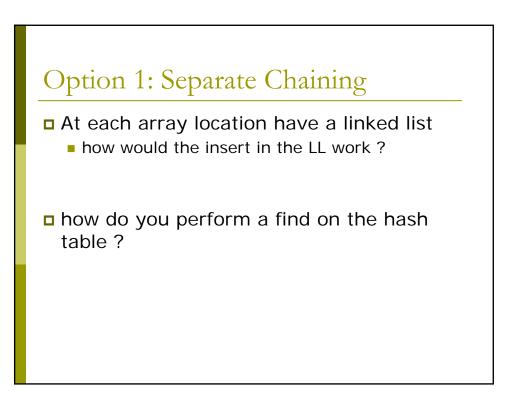
lets add some spread to the summation

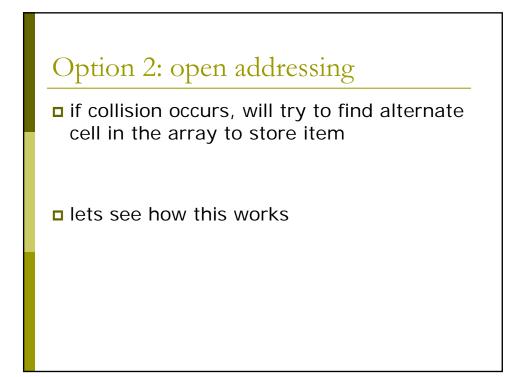
```
Hash(ley) = key[1]*26<sup>0</sup> + key[1]*26<sup>1</sup> *
..key[i]*26<sup>i</sup>
```

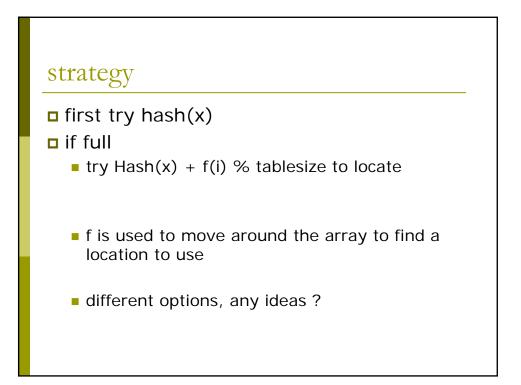














 $\Box f(i) = i$

Example

can you think of any issues ?

clustering

linear probing suffers from a problem called clustering

domino affect

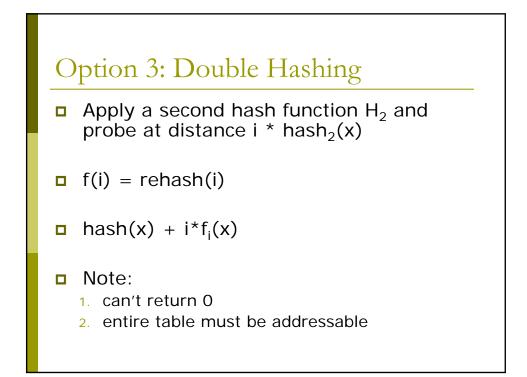
Quadratic probing

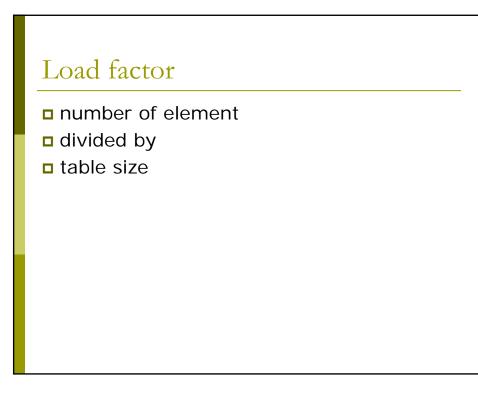
 $\Box f(i) = i^2$

how will this affect clusters ?

Theorem

if quadratic probing is used and table size is prime, and table is at least half empty then we will always find a spot for a new element





growing

So how do you resize a hash ??

deletion

how would deletion work

any issues?

Extendible Hashing

setup similar to B+ tree

hashing routine which has growth built in

use partial bits for keys

when need to grow will use more bits

question

from the data structures we have covered which is the most space efficient ??

