

For Wednesday

- Finish Weiss, chapter 5
- Homework:
 - B+ tree exercise from Blackboard

Paper 1

- Any questions?

External Dictionaries

- We've talked so far about dictionaries small enough to reside in memory
- However, many applications require dictionaries much larger than will easily fit in memory
- Biggest issue for external dictionaries is the number of disk accesses required for an operation
- Each disk access retrieves a block of memory

m-way Search Trees

- Empty tree
or
- Each internal node has up to m children and between 1 and $m-1$ elements
- A node with p elements has exactly $p+1$ children
- Elements are ordered

B-tree of Order m

- an m -way search tree
- If non-empty
 - The root has at least two children
 - All internal nodes other than the root have at least ceiling of $m/2$ children.
 - All external nodes are at the same level
- Thus we have guarantees on the height of the tree
- Book technically covers structure called B+-tree—items in internal nodes also appear in external nodes

Operations

- Searching
- Insertion
- Deletion

Dictionaries

- What's one of our biggest goals in creating a dictionary?

Answer

- Fast search
- How do we do with trees?

New Search Goal

- Constant time search

Hashing

- What's the basic idea?

Perfect Hashing

- What would constitute a perfect hash?

Imperfect Hashing

- Two basic issues:
 - How do we determine the hash bucket from the key?
 - How do we deal with collisions?

Computing the Hash Index

- Typically use mod

Collision Handling Options

- Separate chaining
- Open addressing
 - Linear probing
 - Quadratic probing
 - Double hashing

Separate Chaining

- What is it?
- Advantages?
- Disadvantages?

Open Addressing

- Just going to have the one table, so need to calculate a new location in the table if there's a collision

Linear probing

- How do we find a new location?
- What are advantages of this approach?
- What are potential problems with this approach?

Quadratic Probing

- How do we find the next location with this approach?
- What advantages does it have over linear probing?
- What are possible problems with the approach?

Double Hashing

- What's the approach?
- Advantages?
- Disadvantages?

Notes on Hashing

- Hash buckets
- The importance of primes
- Performance