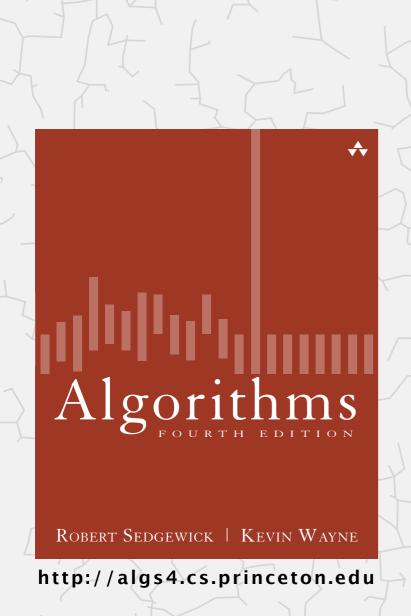
Algorithms



3.1 SYMBOL TABLES

- ▶ API
- elementary implementations
- ordered operations

3.1 SYMBOL TABLES

API

elementary implementations

ordered operations

Algorithms

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

Symbol tables

Key-value pair abstraction.

- Insert a value with specified key.
- Given a key, search for the corresponding value.

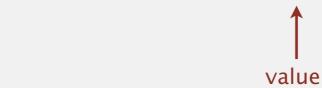
Ex. DNS lookup.

Insert domain name with specified IP address.

key

· Given domain name, find corresponding IP address.

domain name	IP address
www.cs.princeton.edu	128.112.136.11
www.princeton.edu	128.112.128.15
www.yale.edu	130.132.143.21
www.harvard.edu	128.103.060.55
www.simpsons.com	209.052.165.60
<u> </u>	↑



Symbol table applications

application	purpose of search	key	value	
dictionary	find definition	word	definition	
book index	find relevant pages	term	list of page numbers	
file share	find song to download	name of song	computer ID	
financial account	process transactions	account number	transaction details	
web search	find relevant web pages	keyword	list of page names	
compiler	find properties of variables	variable name	type and value	
routing table	route Internet packets	destination	best route	
DNS	find IP address	domain name	IP address	
reverse DNS	find domain name	IP address	domain name	
genomics	find markers	DNA string	known positions	
file system	find file on disk	filename	location on disk	

Symbol tables: context

Also known as: maps, dictionaries, associative arrays.

Generalizes arrays. Keys need not be between 0 and N-1.

Language support.

- External libraries: C, VisualBasic, Standard ML, bash, ...
- Built-in libraries: Java, C#, C++, Scala, ...
- Built-in to language: Python, Perl, PHP, Tcl, JavaScript, Awk, Ruby, Lua.

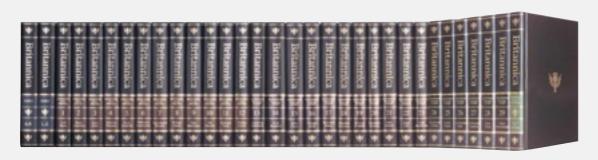
every array is an every object is an table is the only associative array associative array "primitive" data structure

has_nice_syntax_for_associative_arrays["Python"] = True
has_nice_syntax_for_associative_arrays["Java"] = False
legal Python code

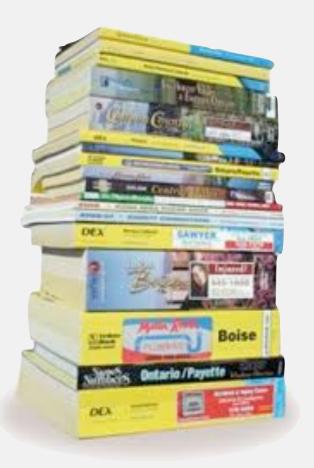
Beyond the basics, or why telephone books are obsolete

Additional operations supported by symbol tables.

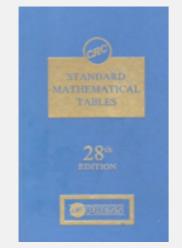
- Change the number associated with a given name.
- · Add a new name, associated with a given number.
- Remove a given name and associated number.



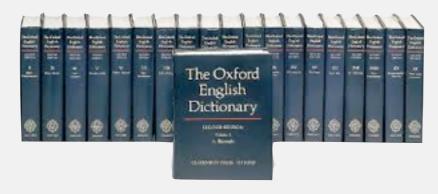
key = term, value = article



key = name value = phone number



key = function name and input
value = function output



key = word, value = definition



key = time and channel value = TV show

Basic symbol table API

Associative array abstraction. Associate one value with each key.

public class	S <mark>ST</mark> <key, value=""></key,>		
	ST()	create an empty symbol table	
void	put(Key key, Value val)	put key-value pair into the table ←	a[key] = val;
Value	get(Key key)	value paired with key	a[key]
boolean	contains(Key key)	is there a value paired with key?	
Iterable <key></key>	keys()	all the keys in the table	syntax in other
void	delete(Key key)	remove key (and its value) from table	languages (not Java)
boolean	isEmpty()	is the table empty?	
int	size()	number of key-value pairs in the table	

Conventions

- Method put() overwrites old value with new value.
- Method get() returns null if key not present.
- Values are not null. ← java.util.Map allows null values

"Careless use of null can cause a staggering variety of bugs.

Studying the Google code base, we found that something like 95% of collections weren't supposed to have any null values in them, and having those fail fast rather than silently accept null would have been helpful to developers."





https://code.google.com/p/guava-libraries/wiki/UsingAndAvoidingNullExplained

Conventions

- Method put() overwrites old value with new value.
- Method get() returns null if key not present.
- Values are not null. ← java.util.Map allows null values

Intended consequences.

• Easy to implement contains().

```
public boolean contains(Key key)
{ return get(key) != null; }
```

• Can implement lazy version of delete().

```
public void delete(Key key)
{  put(key, null); }
```

ST test client for analysis

Frequency counter. Read a sequence of strings from standard input and print out one that occurs with highest frequency.

```
% more tinyTale.txt
it was the best of times
it was the worst of times
it was the age of wisdom
it was the age of foolishness
it was the epoch of belief
it was the epoch of incredulity
it was the season of light
it was the season of darkness
it was the spring of hope
it was the winter of despair
                                                        tiny example
% java FrequencyCounter 3 < tinyTale.txt</pre>
                                                        (60 words, 20 distinct)
the 10
                                                        real example
% java FrequencyCounter 8 < tale.txt</pre>
                                                        (135,635 words, 10,769 distinct)
business 122
                                                        real example
% java FrequencyCounter 10 < leipzig1M.txt ←
                                                        (21,191,455 words, 534,580 distinct)
government 24763
```

Frequency counter implementation

```
public class FrequencyCounter
{
  public static void main(String[] args)
     int minlen = Integer.parseInt(args[0]);
     ST<String, Integer> st = new ST<String, Integer>();
                                                                     create ST
     while (!StdIn.isEmpty())
        if (word.length() < minlen) continue;</pre>
                                                                     read string and
        if (!st.contains(word)) st.put(word, 1);
                                                                     update frequency
                                st.put(word, st.get(word) + 1);
        else
     String max = "";
                                       print a string with max frequency
     st.put(max, 0);
     for (String word : st.keys())
        if (st.get(word) > st.get(max))
           max = word;
     StdOut.println(max + " " + st.get(max));
}
```

Keys and values

Value type. Any generic type.

- Best practice: use immutable types for symbol table keys.
 - Immutable in Java: Integer, Double, String, java.io.File, ...
 - Mutable in Java: StringBuilder, java.net.URL, arrays, ...

specify Comparable in API.

Key type: several natural assumptions.

- Assume keys are Comparable, use compareTo().
- Assume keys are any generic type, use equals() to test equality.
- Assume keys are any generic type, use equals() to test equality;
 use hashCode() to scramble key.

built-in to Java (stay tuned)

Equality test

All Java classes inherit a method equals().

Java requirements. For any references x, y and z:

- Reflexive: x.equals(x) is true.
- Symmetric: x.equals(y) iff y.equals(x).
- Transitive: if x.equals(y) and y.equals(z), then x.equals(z).
- Non-null: x.equals(null) is false.

```
do x and y refer to
the same object?
```

Default implementation. (x == y)

Customized implementations. Integer, Double, String, java.io.File, ...

User-defined implementations. Some care needed.

13

Implementing equals for user-defined types

Seems easy.

```
public
       class Date implements Comparable<Date>
   private final int month;
   private final int day;
   private final int year;
   public boolean equals(Date that)
                                                          check that all significant
      if (this.day != that.day ) return false;
                                                          fields are the same
      if (this.month != that.month) return false;
      if (this.year != that.year ) return false;
      return true;
```

Implementing equals for user-defined types

typically unsafe to use equals() with inheritance Seems easy, but requires some care. (would violate symmetry) public final class Date implements Comparable<Date> private final int month; must be Object. private final int day; Why? Experts still debate. private final int year; public boolean equals(Object y) optimize for true object equality if (y == this) return true; check for null if (y == null) return false; objects must be in the same class if (y.getClass() != this.getClass()) (could also use instanceof) return false; Date that = (Date) y; cast is guaranteed to succeed if (this.day != that.day) return false; check that all significant if (this.month != that.month) return false; fields are the same if (this.year != that.year) return false; return true;

Equals design

"Standard" recipe for user-defined types.

- Optimization for reference equality.
- Check against null.
- Check that two objects are of the same type; cast.
- Compare each significant field:

 - if field is an object, use equals()
 ← apply rule recursively
 - if field is an array, apply to each entry ← can use Arrays.deepEquals(a, b)
 but not a.equals(b)

Best practices.

- e.g., cached Manhattan distance
- No need to use calculated fields that depend on other fields.
- Compare fields mostly likely to differ first.
- Make compareTo() consistent with equals().

```
x.equals(y) if and only if (x.compareTo(y) == 0)
```

3.1 SYMBOL TABLES

APH

- elementary implementations
- ordered operations

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

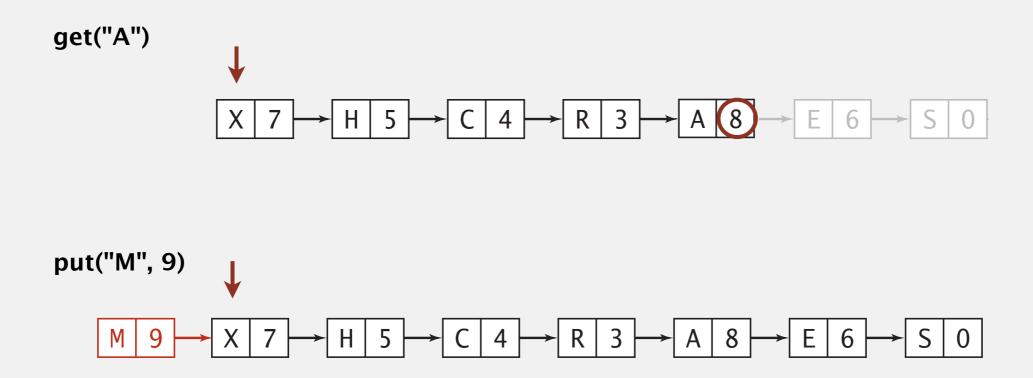
http://algs4.cs.princeton.edu

Sequential search in a linked list

Data structure. Maintain an (unordered) linked list of key-value pairs.

Search. Scan through all keys until match found.

Insert. Scan through all keys until match found; if no match, add to front.



Elementary ST implementations: summary

implementation	guarantee		average case		operations
implementation	search	insert	search hit	insert	on keys
sequential search (unordered list)	N	N	N	N	equals()

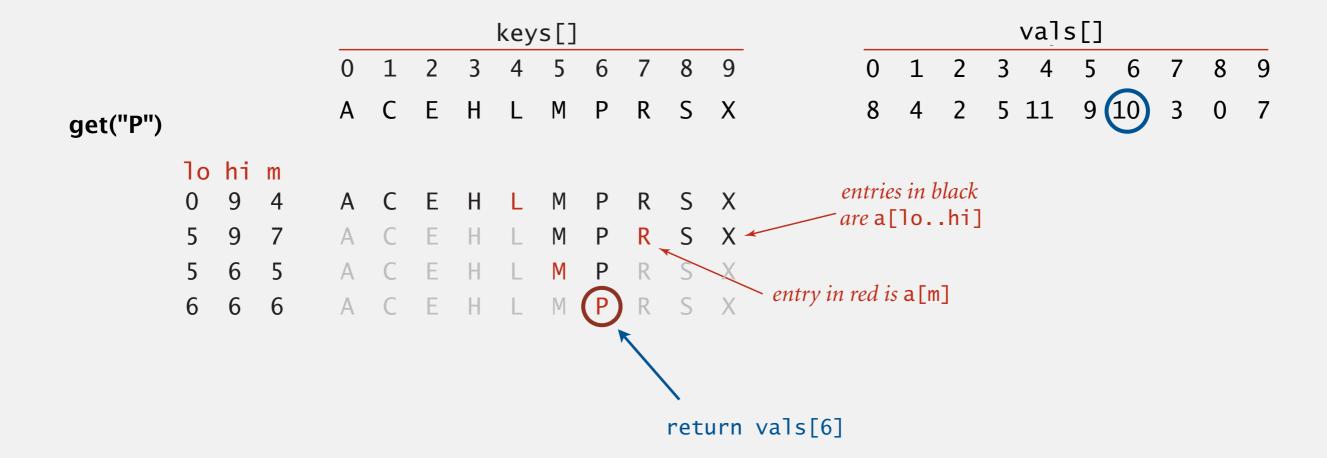
Challenge. Efficient implementations of both search and insert.

Binary search in an ordered array

Data structure. Maintain parallel arrays for keys and values, sorted by keys.

Search. Use binary search to find key.

Proposition. At most $\sim \lg N$ compares to search a sorted array of length N.



Binary search in an ordered array

Data structure. Maintain parallel arrays for keys and values, sorted by keys.

Search. Use binary search to find key.

```
public Value get(Key key)
   int lo = 0, hi = N-1;
   while (lo <= hi)
    {
       int mid = lo + (hi - lo) / 2;
        int cmp = key.compareTo(keys[mid]);
        if (cmp < 0) hi = mid - 1;
        else if (cmp > 0) lo = mid + 1;
        else if (cmp == 0) return vals[mid];
   return null; ← no matching key
}
```

Binary search: insert

Data structure. Maintain parallel arrays for keys and values, sorted by keys.

Search. Use binary search to find key.

Insert. Use binary search to find place to insert; shift all larger keys over. Proposition. Takes linear time in the worst case.

put("P", 10)

keys[]
2 3 4 5 6 7 8 9
S X

Elementary ST implementations: summary

implementation	guarantee		average case		operations
implementation	search	insert	search hit	insert	on keys
sequential search (unordered list)	N	N	N	N	equals()
binary search (ordered array)	log N	N^{\dagger}	log N	N^{\dagger}	compareTo()

† can do with log N compares, but requires N array accesses

Challenge. Efficient implementations of both search and insert.

3.1 SYMBOL TABLES

API

- elementary implementations
- ordered operations

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

Examples of ordered symbol table API

```
values
                                keys
                    min() \rightarrow 09:00:00 Chicago
                             09:00:03 Phoenix
                             09:00:13→ Houston
            get(09:00:13) 09:00:59 Chicago
                             09:01:10
                                         Houston
          floor(09:05:00) \rightarrow 09:03:13
                                         Chicago
                                         Seattle
                             09:10:11
                select(7) \longrightarrow 09:10:25 Seattle
                             09:14:25 Phoenix
                             09:19:32
                                         Chicago
                             09:19:46
                                         Chicago
keys(09:15:00, 09:25:00) \longrightarrow 09:21:05
                                        Chicago
                             09:22:43
                                         Seattle
                             09:22:54 Seattle
                             09:25:52
                                         Chicago
       ceiling(09:30:00) \rightarrow 09:35:21
                                         Chicago
                                         Seattle
                             09:36:14
                    max() \longrightarrow 09:37:44 Phoenix
size(09:15:00, 09:25:00) is 5
     rank(09:10:25) is 7
```

Ordered symbol table API

```
public class ST<Key (extends Comparable<Key>,) Value>
Key min()
                                            smallest key
Key max()
                                            largest key
Key floor(Key key)
                                 largest key less than or equal to key
Key ceiling(Key key)
                               smallest key greater than or equal to key
int rank(Key key)
                                     number of keys less than key
Key select(int k)
                                           key of rank k
```

RANK IN A SORTED ARRAY

Problem. Given a sorted array of *N* distinct keys, find the number of keys strictly less than a given query key.

RANK IN A SORTED ARRAY

Problem. Given a sorted array of *N* distinct keys, find the number of keys strictly less than a given query key.

easy modification to binary search

```
public Value get(Key key) public int rank(Key key)
   int lo = 0, hi = N-1;
   while (lo <= hi)
       int mid = 10 + (hi - 10) / 2;
       int cmp = key.compareTo(keys[mid]);
       if (cmp < 0) hi = mid - 1;
       else if (cmp > 0) lo = mid + 1;
       else if (cmp == 0) return vals[mid]; mid
   return null; lo
}
```

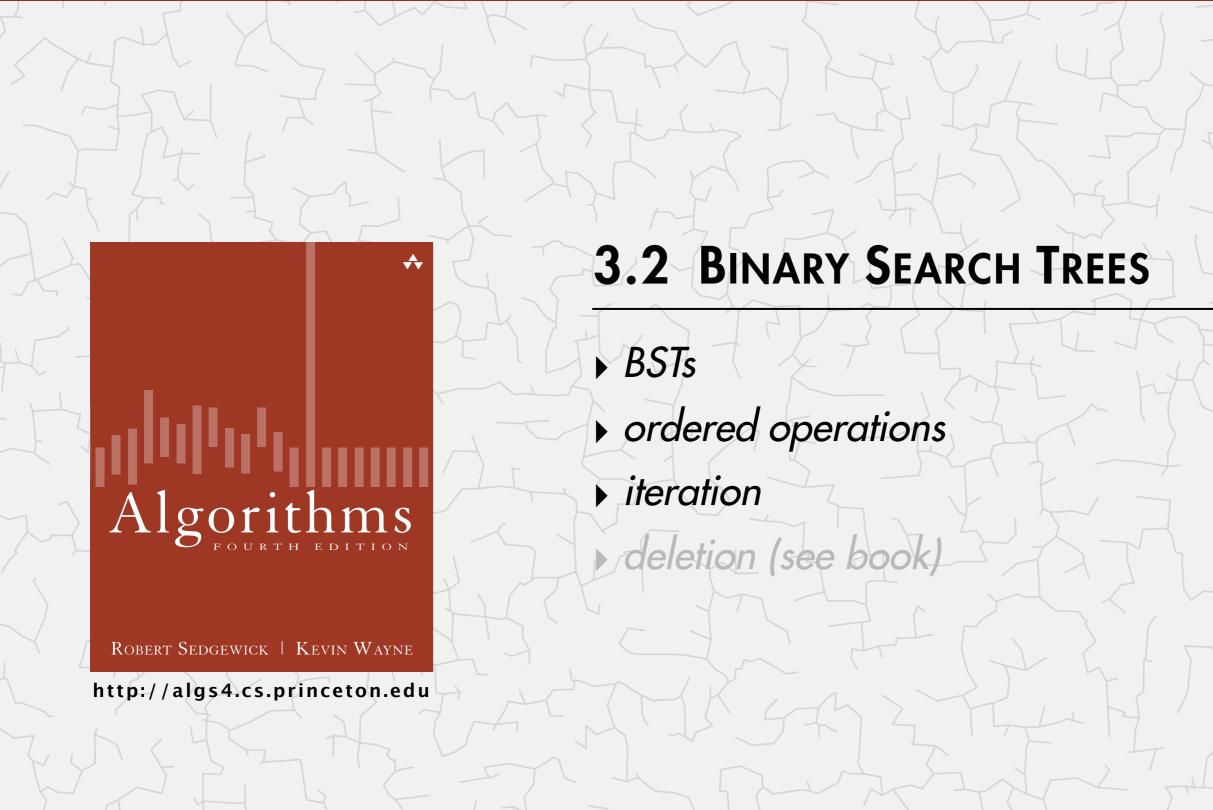
Binary search: ordered symbol table operations summary

	sequential search	binary search
search	N	$\log N$
insert	N	N
min / max	N	1
floor / ceiling	N	$\log N$
rank	N	$\log N$
select	N	1

order of growth of the running time for ordered symbol table operations

Last updated on Oct 9, 2015, 11:22 AM

Algorithms



Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

3.2 BINARY SEARCH TREES

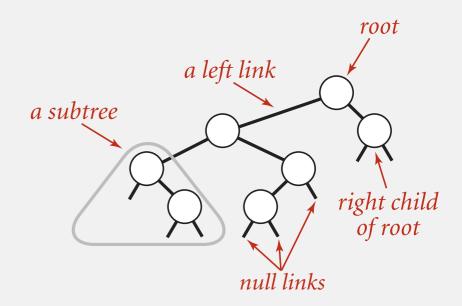
- ▶ BSTs
- ordered operations
- iteration
- deletion

Binary search trees

Definition. A BST is a binary tree in symmetric order.

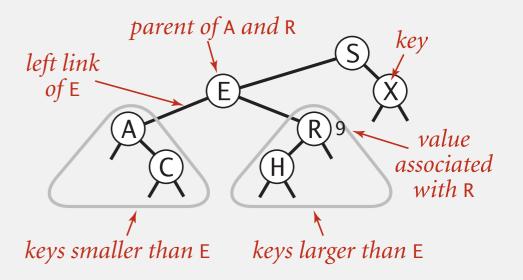
A binary tree is either:

- Empty.
- Two disjoint binary trees (left and right).



Symmetric order. Each node has a distinct key, and every node's key is:

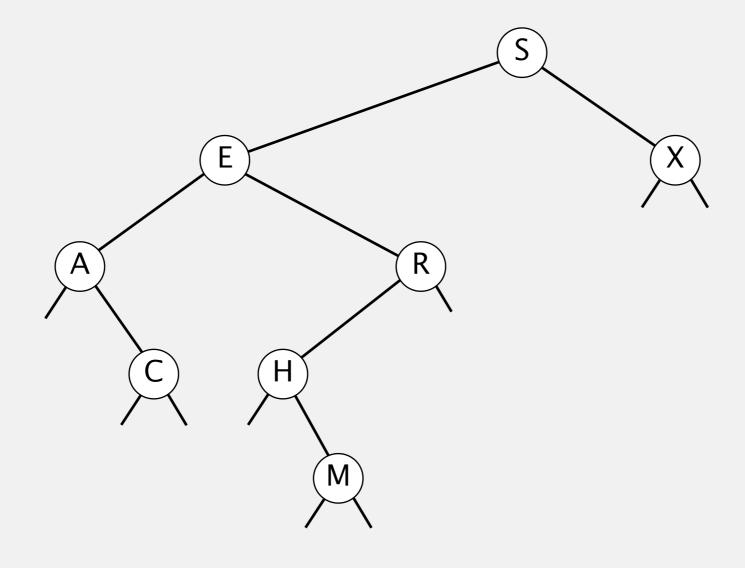
- Larger than all keys in its left subtree.
- Smaller than all keys in its right subtree.



Binary search tree demo

Search. If less, go left; if greater, go right; if equal, search hit.

successful search for H

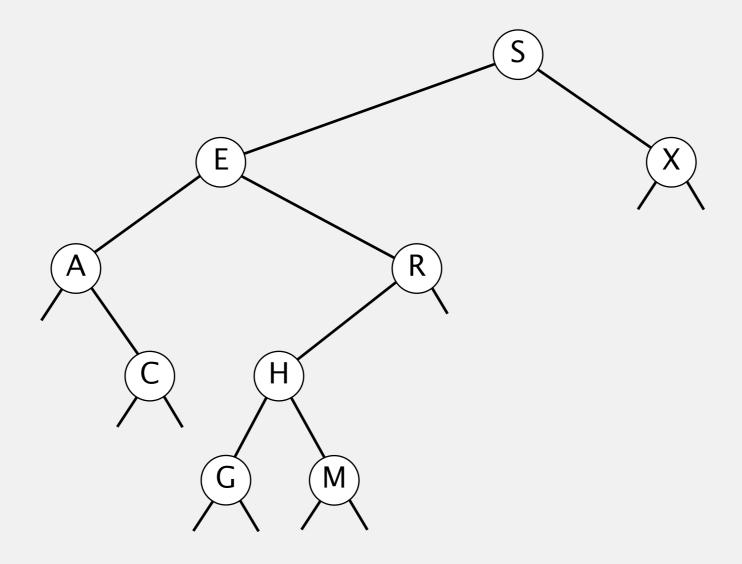




Binary search tree demo

Insert. If less, go left; if greater, go right; if null, insert.

insert G





BST representation in Java

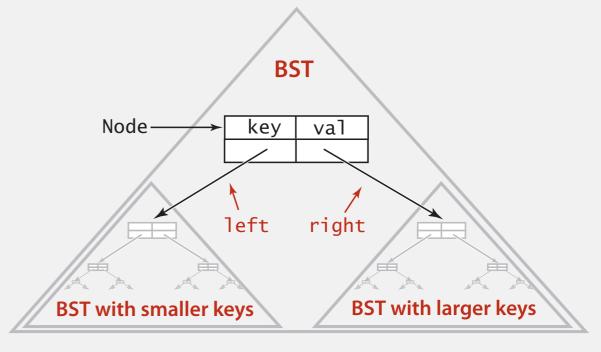
Java definition. A BST is a reference to a root Node.

A Node is composed of four fields:

- A Key and a Value.
- A reference to the left and right subtree.

```
smaller keys larger keys
```

```
private class Node
{
   private Key key;
   private Value val;
   private Node left, right;
   public Node(Key key, Value val)
   {
      this.key = key;
      this.val = val;
   }
}
```



Binary search tree

Key and Value are generic types; Key is Comparable

BST implementation (skeleton)

```
public class BST<Key extends Comparable<Key>, Value>
   private Node root;
                                                           root of BST
  private class Node
  { /* see previous slide */ }
  public void put(Key key, Value val)
   { /* see next slides */ }
  public Value get(Key key)
   { /* see next slides */ }
  public Iterable<Key> iterator()
   { /* see slides in next section */ }
  public void delete(Key key)
   { /* see textbook */ }
```

BST search: Java implementation

Get. Return value corresponding to given key, or null if no such key.

```
public Value get(Key key)
{
   Node x = root;
   while (x != null)
   {
      int cmp = key.compareTo(x.key);
      if (cmp < 0) x = x.left;
      else if (cmp > 0) x = x.right;
      else if (cmp == 0) return x.val;
   }
   return null;
}
```

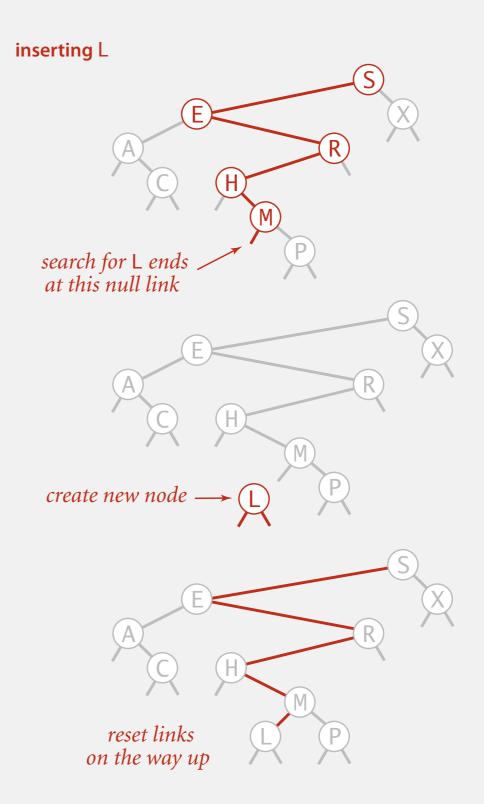
Cost. Number of compares = 1 + depth of node.

BST insert

Put. Associate value with key.

Search for key, then two cases:

- Key in tree ⇒ reset value.
- Key not in tree ⇒ add new node.



Insertion into a BST

BST insert: Java implementation

Put. Associate value with key.

```
public void put(Key key, Value val)
{    root = put(root, key, val); }

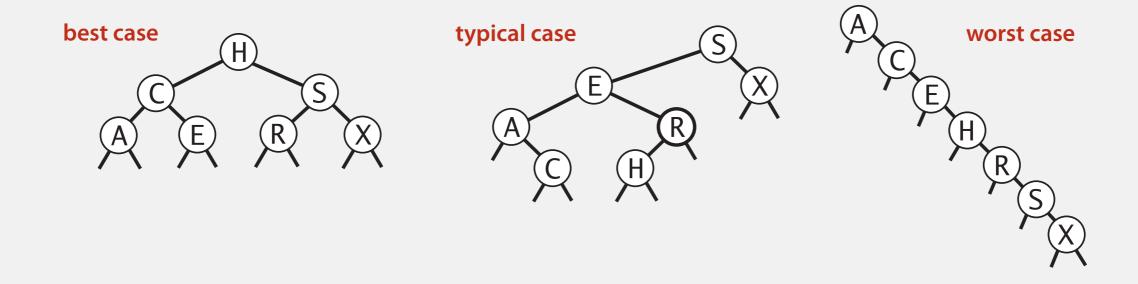
private Node put(Node x, Key key, Value val)
{
    if (x == null) return new Node(key, val);
    int cmp = key.compareTo(x.key);
    if (cmp < 0) x.left = put(x.left, key, val);
    else if (cmp > 0) x.right = put(x.right, key, val);
    else if (cmp == 0) x.val = val;
    return x;
}

Marning: concise but tricky code; read carefully!
```

Cost. Number of compares = 1 + depth of node.

Tree shape

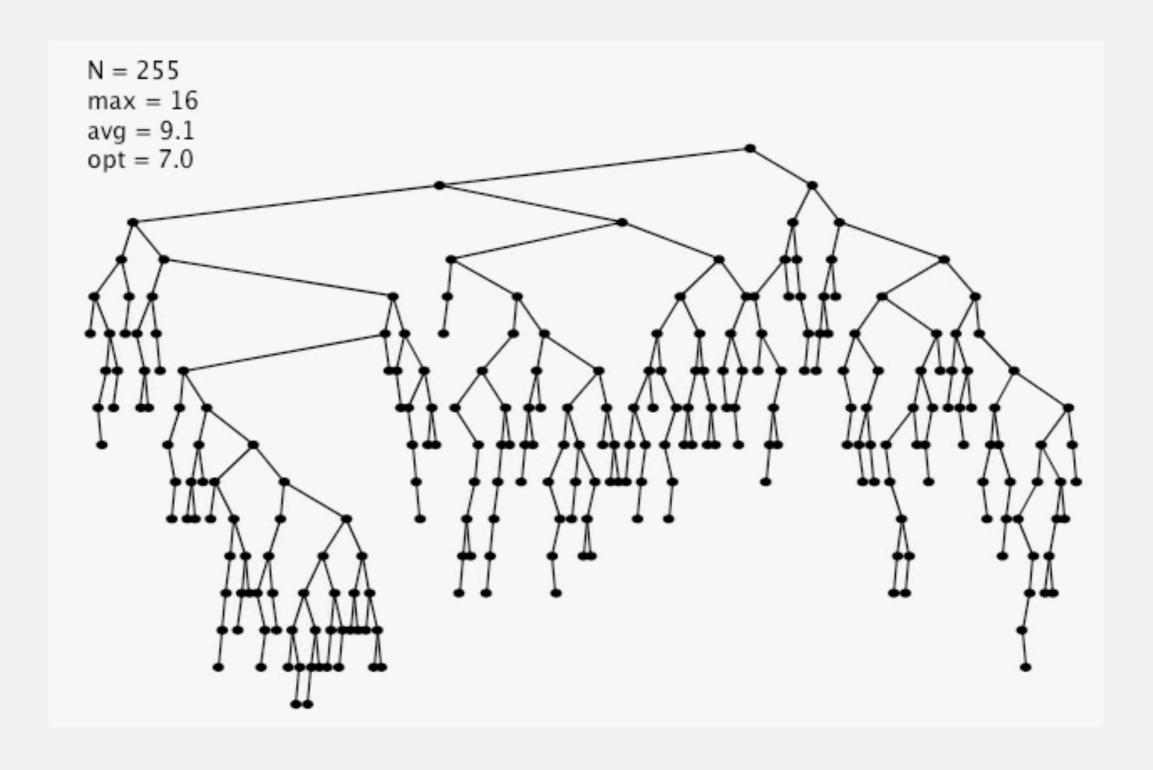
- Many BSTs correspond to same set of keys.
- Number of compares for search/insert = 1 + depth of node.



Bottom line. Tree shape depends on order of insertion.

BST insertion: random order visualization

Ex. Insert keys in random order.



3.2 BINARY SEARCH TREES

- BSTs
- iteration
- ordered operations
- deletion

Robert Sedgewick | Kevin Wayne

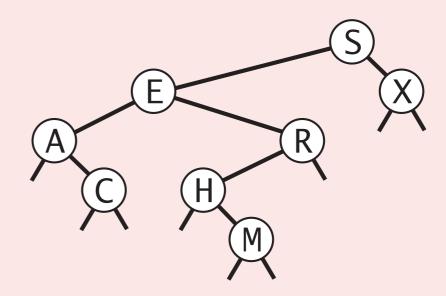
Algorithms

http://algs4.cs.princeton.edu

Binary search trees: quiz 1

In what order does the traverse(root) code print out the keys in the BST?

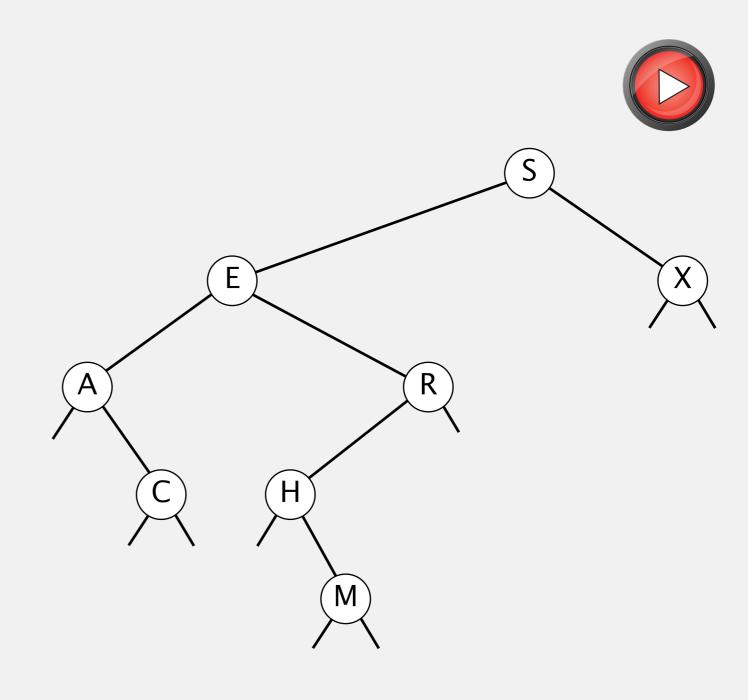
```
private void traverse(Node x)
{
   if (x == null) return;
   traverse(x.left);
   StdOut.println(x.key);
   traverse(x.right);
}
```



- A. ACEHMRSX
- B. ACERHMXS
- C. SEACRHMX
- D. CAMHREXS
- **E.** *I don't know.*

Inorder traversal

```
inorder(S)
   inorder(E)
      inorder(A)
         print A
         inorder(C)
            print C
            done C
         done A
      print E
      inorder(R)
         inorder(H)
            print H
            inorder(M)
               print M
               done M
            done H
         print R
         done R
      done E
   print S
   inorder(X)
      print X
      done X
   done S
```



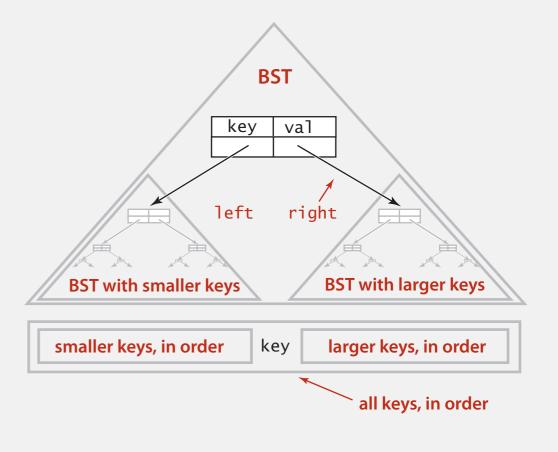
output: A C E H M R S X

Inorder traversal

- Traverse left subtree.
- Enqueue key.
- Traverse right subtree.

```
public Iterable<Key> keys()
{
    Queue<Key> q = new Queue<Key>();
    inorder(root, q);
    return q;
}

private void inorder(Node x, Queue<Key> q)
{
    if (x == null) return;
    inorder(x.left, q);
    q.enqueue(x.key);
    inorder(x.right, q);
}
```



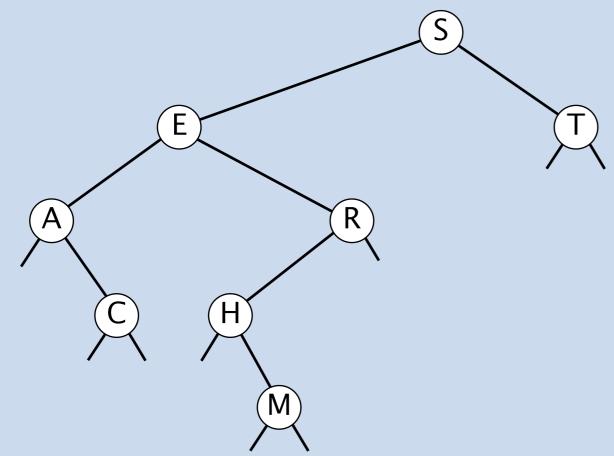
Property. Inorder traversal of a BST yields keys in ascending order.

LEVEL-ORDER TRAVERSAL

Level-order traversal of a binary tree.

- Process root.
- Process children of root, from left to right.
- · Process grandchildren of root, from left to right.

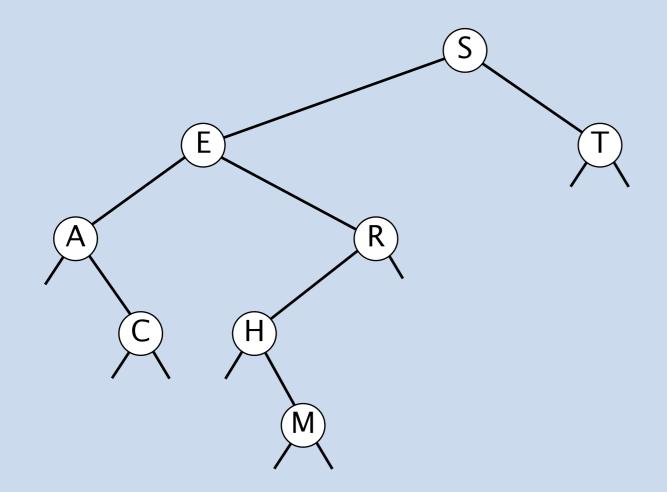
• ...



level order traversal: SETARCHM

LEVEL-ORDER TRAVERSAL

Q1. Given binary tree, how to compute level-order traversal?

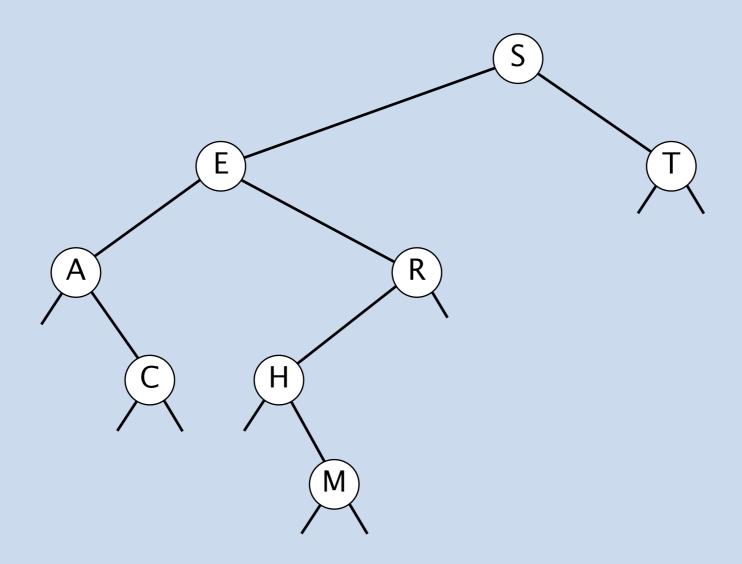


```
queue.enqueue(root);
while (!queue.isEmpty())
{
    Node x = queue.dequeue();
    if (x == null) continue;
    StdOut.println(x.item);
    queue.enqueue(x.left);
    queue.enqueue(x.right);
}
```

level order traversal: SETARCHM

LEVEL-ORDER TRAVERSAL

Q2. Given level-order traversal of a BST, how to (uniquely) reconstruct BST?



Binary search trees: quiz 2

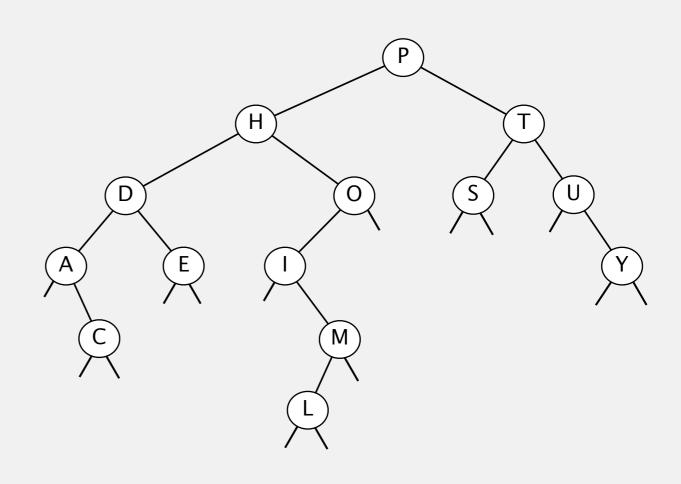
Given distinct keys, to which sorting algorithm is this conceptually closest?

- 1. Shuffle the keys.
- 2. Insert the keys into a BST, one at a time.
- 3. Do an inorder traversal of the BST.

- A. Insertion sort.
- **B.** Mergesort.
- C. Quicksort.
- **D.** *None of the above.*
- **E.** *I don't know.*

Correspondence between BSTs and quicksort partitioning





Remark. Correspondence is 1–1 if array has no duplicate keys.

BSTs: mathematical analysis

Proposition. If N distinct keys are inserted into a BST in random order, the expected number of compares for a search/insert is $\sim 2 \ln N$. Pf. 1–1 correspondence with quicksort partitioning.

Proposition. [Reed, 2003] If N distinct keys are inserted into a BST in random order, the expected height is $\sim 4.311 \ln N$.

expected depth of function-call stack in quicksort

How Tall is a Tree?

Bruce Reed CNRS, Paris, France reed@moka.ccr.jussieu.fr

ABSTRACT

Let H_n be the height of a random binary search tree on n nodes. We show that there exists constants $\alpha = 4.31107...$ and $\beta = 1.95...$ such that $\mathbf{E}(H_n) = \alpha \log n - \beta \log \log n + O(1)$, We also show that $Var(H_n) = O(1)$.

But... Worst-case height is N-1.

[exponentially small chance when keys are inserted in random order]

ST implementations: summary

implementation	guarantee		average case		operations			
	search	insert	search hit	insert	on keys			
sequential search (unordered list)	N	N	N	N	equals()			
binary search (ordered array)	log N	N	log N	N	compareTo()			
BST	N	N	log N	log N	compareTo()			

Why not shuffle to ensure a (probabilistic) guarantee of log N?

3.2 BINARY SEARCH TREES BSTs

- iteration
- ordered operations
- deletion

Algorithms

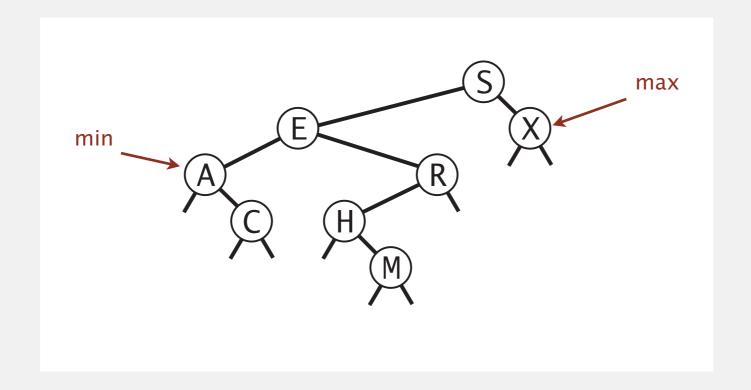
ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

Minimum and maximum

Minimum. Smallest key in BST.

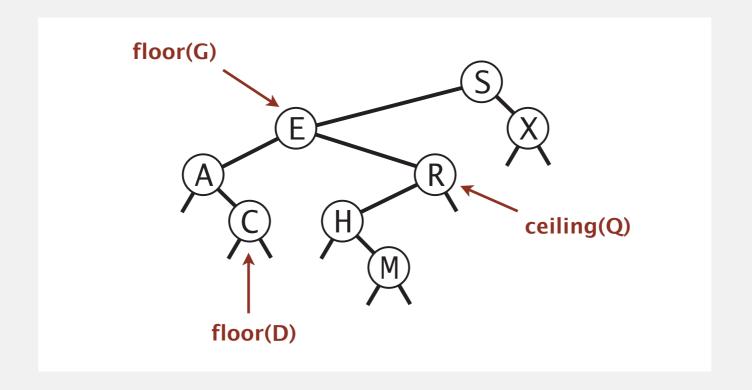
Maximum. Largest key in BST.



Q. How to find the min / max?

Floor and ceiling

Floor. Largest key in BST ≤ query key. Ceiling. Smallest key in BST ≥ query key.



Q. How to find the floor / ceiling?



Computing the floor

Floor. Largest key in BST $\leq k$?

Case 1. [key in node x = k]

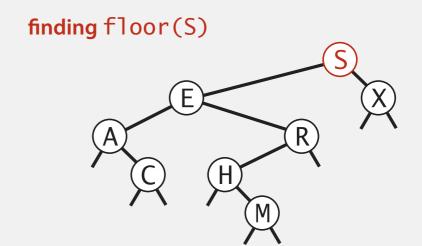
The floor of k is k.

Case 2. [key in node x > k]

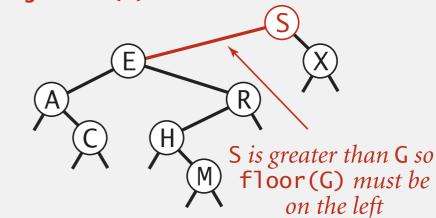
The floor of k is in the left subtree of x.

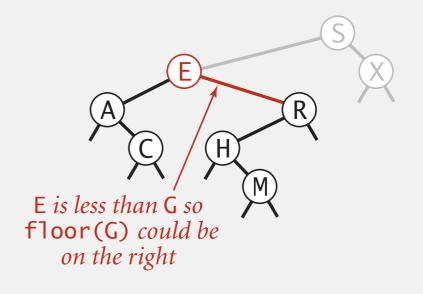
Case 3. [key in node x < k]

The floor of k can't be in left subtree of x: it is either in the right subtree of x or it is the key in node x.



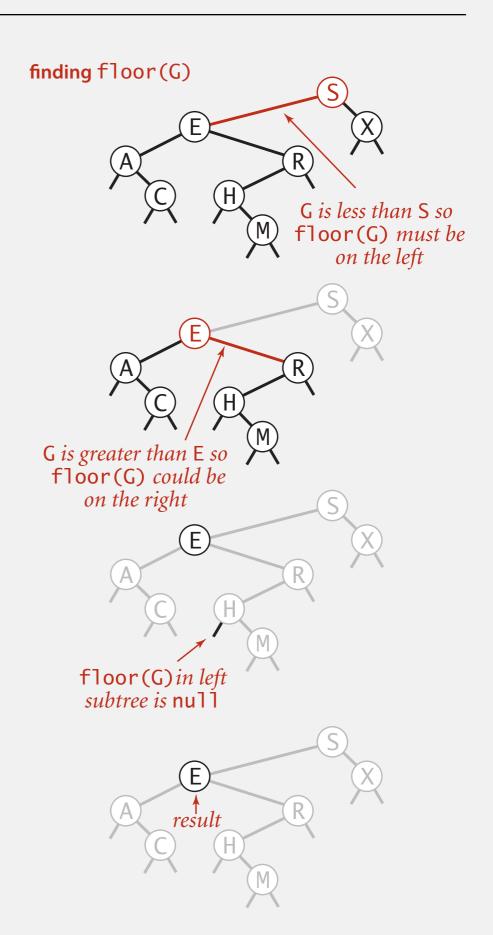






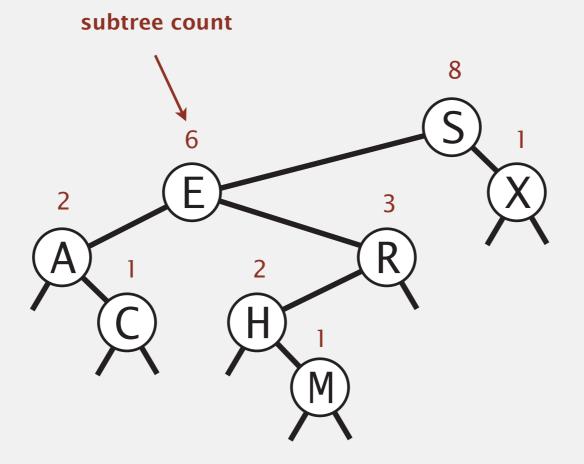
Computing the floor

```
public Key floor(Key key)
  return floor(root, key); }
private Key floor(Node x, Key key)
   if (x == null) return null;
   int cmp = key.compareTo(x.key);
   if (cmp == 0) return x;
   if (cmp < 0) return floor(x.left, key);</pre>
   Key t = floor(x.right, key);
   if (t != null) return t;
                  return x.key;
   else
```



Rank and select

- Q. How to implement rank() and select() efficiently for BSTs?
- A. In each node, store the number of nodes in its subtree.



BST implementation: subtree counts

```
private Node put(Node x, Key key, Value val)

initialize subtree
count to l

if (x == null) return new Node(key, val, 1);
int cmp = key.compareTo(x.key);
if (cmp < 0) x.left = put(x.left, key, val);
else if (cmp > 0) x.right = put(x.right, key, val);
else if (cmp == 0) x.val = val;
x.count = 1 + size(x.left) + size(x.right);
return x;
}
```

Computing the rank

Rank. How many keys in BST < k?

Case 1. [k < key in node]

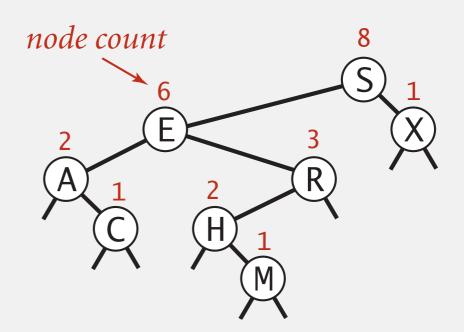
- Keys in left subtree? count
- Key in node?
- Keys in right subtree?

Case 2. [k > key in node]

- Keys in left subtree? all
- Key in node.
- Keys in right subtree? *count*

Case 3. [k = key in node]

- Keys in left subtree? count
- Key in node.
- Keys in right subtree? 0

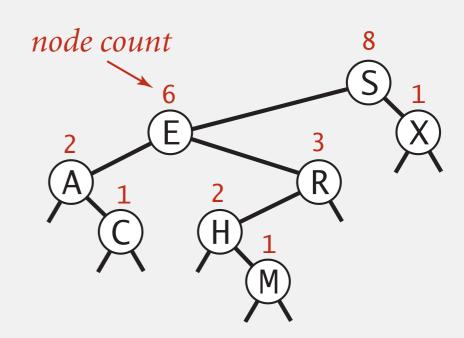




Rank

Rank. How many keys in BST < k?

Easy recursive algorithm (3 cases!)



BST: ordered symbol table operations summary

	sequential search	binary search	BST	
search	N	log N	h	
insert	N	N	h	h = height of BST
min / max	N	1	h	(proportional to log N if keys inserted in random order)
floor / ceiling	N	log N	h	
rank	N	log N	h	
select	N	1	h	
ordered iteration	$N \log N$	N	N	

order of growth of running time of ordered symbol table operations

ST implementations: summary

implementation	guarantee		average case		ordered	key
	search	insert	search hit	insert	ops?	interface
sequential search (unordered list)	N	N	N	N		equals()
binary search (ordered array)	log N	N	log N	N	✓	compareTo()
BST	N	N	log N	log N	•	compareTo()
red-black BST	$\log N$	$\log N$	log N	log N	•	compareTo()

Next lecture. Guarantee logarithmic performance for all operations.