Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

Algorithms

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

3.1 SYMBOL TABLES

elementary implementations

ordered operations

► API

"Smart data structures and dumb code works a lot better than the other way around." – Eric S. Raymond



WITH A FOREWORD BY BOB YOUNG, CHAIRMAN & CEO OF RED HAT, INC.

3.1 SYMBOL TABLES

ordered operations

elementary implementations

API

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.
 - 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
↑ key	value

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	

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: Awk, Perl, PHP, Tcl, JavaScript, Python, Ruby, Lua.



hasNiceSyntaxForAssociativeArrays["Python"] = True hasNiceSyntaxForAssociativeArrays["Java"] = False

legal Python code

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 \leftarrow	<pre>_ a[key] = val;</pre>
Value	get(Key key)	value paired with key \checkmark	_ a[key]
boolean	contains(Key key)	is there a value paired with key?	
Iterable <key></key>	keys()	all the keys in the table	
void	delete(Key key)	remove key (and its value) from table	
boolean	isEmpty()	is the table empty?	
int	size()	number of key-value pairs in the table	

Conventions

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

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); }
```

Value type. Any generic type.

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.

Best practices. Use immutable types for symbol table keys.

built-in to Java

(stay tuned)

- Immutable in Java: Integer, Double, String, java.io.File, ...
- Mutable in Java: StringBuilder, java.net.URL, arrays, ...

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.

equivalence

relation

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



Equals design

Best practices.

"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 a primitive type, use ==
 - if field is an object, use equals()
 - if field is an array, apply to each entry

____ but use Double.compare() with double (to deal with -0.0 and NaN)

- apply rule recursively
- can use Arrays.deepEquals(a, b)
 but not a.equals(b)

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)

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



Frequency counter implementation



3.1 SYMBOL TABLES

ordered operations

API

elementary implementations

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

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

Search. Scan through all keys until find a match.

Insert. Scan through all keys until find a match; if no match add to front.



implomentation	guara	antee	averag	average case operations search hit insert		
Implementation	search	insert	search hit	insert	on keys	
sequential search (unordered list)	Ν	Ν	Ν	Ν	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*.



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

Search. Use binary search to find key.

Implementing binary search was

- A. Easier than I thought.
- **B.** About what I expected.
- C. Harder than I thought.
- **D.** Much harder than I thought.
- E. I don't know.

Problem. Given an array with all 0s in the beginning and all 1s at the end, find the index in the array where the 1s begin.

Variant 1. You are given the length of the array.Variant 2. You are not given the length of the array.

Data structure. Maintain an ordered array of key-value pairs.

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)



implomentation	guara	antee	average case		operations	
implementation	search	insert	search hit	insert	on keys	
sequential search (unordered list)	Ν	Ν	Ν	Ν	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

elementary implementations

Algorithms

ordered operations

API

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

	keys	values
min()→	►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)→	09:03:13	Chicago
	09:10:11	Seattle
select(7)→	►09:10:25	Seattle
	09:14:25	Phoenix
	09:19:32	Chicago
	09:19:46	Chicago
keys(09:15:00, 09:25:00)→	09:21:05	Chicago
	09:22:43	Seattle
	09:22:54	Seattle
	09:25:52	Chicago
ceiling(09:30:00) →	09:35:21	Chicago
	09:36:14	Seattle
max() —	09:37:44	Phoenix
size(09:15:00, 09:25:00)	5	

publ	public class ST <key comparable<key="" extends="">, Value></key>				
	• •				
Key	min()	smallest key			
Кеу	max()	largest key			
Кеу	floor(Key key)	largest key less than or equal to key			
Кеу	<pre>ceiling(Key key)</pre>	smallest key greater than or equal to key			
int	rank(Key key)	number of keys less than key			
Кеу	<pre>select(int k)</pre>	key of rank k			
	÷				

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

Binary search: ordered symbol table operations summary

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

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

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE



3.2 BINARY SEARCH TREES

► BSTs

iteration

deletion

ordered operations

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

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 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





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





Binary search tree

Key and Value are generic types; Key is Comparable

BST implementation (skeleton)



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.

Put. Associate value with key.

Search for key, then two cases:

- Key in tree \Rightarrow reset value.
- Key not in tree \Rightarrow add new node.



Insertion into a BST

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;
}
Warning: 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.



Binary search trees: quiz 1

Given *N* distinct keys, what is the name of this sorting algorithm?

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.

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.



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

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

implementation	guara	antee	averag	average case		
	search	insert	search hit	insert	on keys	
sequential search (unordered list)	Ν	Ν	Ν	Ν	equals()	
binary search (ordered array)	log N	Ν	log N	Ν	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

iteration

deletion

ordered operations

BSTs

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

http://algs4.cs.princeton.edu

Binary search trees: quiz 2

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. A C E H M R S X
- B. A C E R H M X S
- 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



- 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.dequeue(x.right);

}

level order traversal: SETARCHM

LEVEL-ORDER TRAVERSAL

- Q2. Given level-order traversal of a BST, how to (uniquely) reconstruct BST?
- Ex. \$ \$7 A \$ \$ U H



3.2 BINARY SEARCH TREES

Algorithms

ordered operations

BSTs

iteration

deletion

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. Largest key in BST \leq query key. Ceiling. Smallest key in BST \geq query key.



Q. How to find the floor / ceiling?



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

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;
else return x.key;

}



- 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



30

Rank. How many keys in BST < *k*?

Case 1. [k < key in node] No key in right subtree < k; some keys in left subtree < k.

Case 2. [k > key in node] All keys in left subtree < k; the key in the node is < k; some keys in right subtree may be < k.

Case 3. [k = key in node] All keys in left subtree < k; no key in right subtree < k.





Rank

Rank. How many keys in BST < *k*?

Easy recursive algorithm (3 cases!)



```
public int rank(Key key)
{ return rank(key, root); }
private int rank(Key key, Node x)
{
    if (x == null) return 0;
    int cmp = key.compareTo(x.key);
    if (cmp < 0) return rank(key, x.left);
    else if (cmp > 0) return 1 + size(x.left) + rank(key, x.right);
    else if (cmp == 0) return size(x.left);
```



order of growth of running time of ordered symbol table operations

implementation	guara	antee	average case ordered		key	
	search	insert	search hit	insert	ops?	interface
sequential search (unordered list)	Ν	N	N	N		equals()
binary search (ordered array)	log N	N	log N	Ν	~	compareTo()
BST	Ν	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.