Symbol Tables

CS 121: Data Structures

START RECORDING

Outline

- Attendance quiz
- Symbol Table API
- Interfaces: Comparable and Equals
- Elementary implementations
- Ordered operations

Attendance Quiz

Attendance Quiz: Iterators and Exceptions

- Scan the QR code, or find today's attendance quiz under the "Quizzes" tab on Canvas
- Password: to be announced in class

Attendance Quiz: Iterators and Exceptions

- Write your name and the date
- Describe one benefit of using the for-each loop (i.e., the iterator syntax) instead of a conventional for-loop
- Use the for-each loop to print the elements in the array:
  `String[] alpha = {"A", "B", "C", …};`
- In a few sentences, explain the importance of exception handling

http://algs4.cs.princeton.edu

3.1 SYMBOL TABLES

- *API*
- *elementary implementations*
- *ordered operations*

http://algs4.cs.princeton.edu

3.1 Symbol Tables

- *API*
- *elementary implementations*
- *ordered operations*

Symbol tables

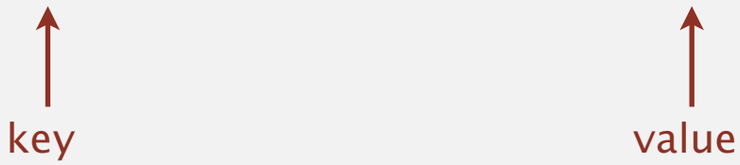
Key-value pair abstraction.

- 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.clarku.edu	140.232.1.129
www.princeton.edu	128.112.128.15
www.google.com	142.250.64.110
www.facebook.com	31.13.71.36
www.simpsons.com	209.052.165.60



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

every array is an associative array

every object is an associative array

table is the only primitive data structure

```
hasNiceSyntaxForAssociativeArrays["Python"] = true
hasNiceSyntaxForAssociativeArrays["Java"]   = false
```

legal Python code

Basic symbol table API

Associative array abstraction. Associate one value with each key.

public class ST<Key, Value>		
	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?*
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*
Iterable<Key>	keys()	*all the keys in the table*

Conventions

- Values are not `null`. ← Java allows null value
- 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);  }
```

Keys and values

Value type. Any generic type.

Key type: several natural assumptions.

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

specify `Comparable` in API.

built-in to Java
(stay tuned)

Best practices. Use immutable types for symbol table keys.

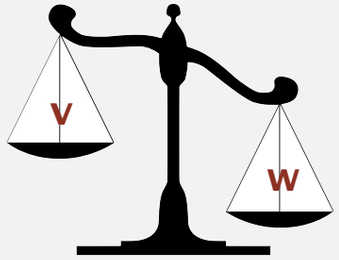
- Immutable in Java: `Integer`, `Double`, `String`, `java.io.File`, ...
- Mutable in Java: `StringBuilder`, `java.net.URL`, arrays, ...

Comparable

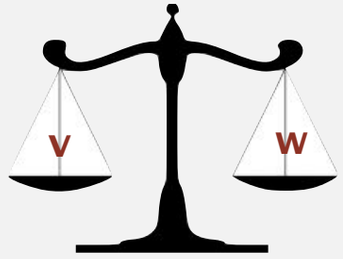
Comparable API

Implement `compareTo()` so that `v.compareTo(w)`

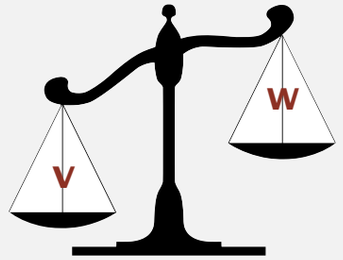
- Defines a total order.
- Returns a negative integer, zero, or positive integer
  if v is less than, equal to, or greater than w, respectively.
- Throws an exception if incompatible types (or either is `null`).



less than (return -1)



equal to (return 0)



greater than (return +1)

Built-in comparable types. `Integer`, `Double`, `String`, `Date`, `File`, ...

User-defined comparable types. Implement the `Comparable` interface.

Implementing the Comparable interface

Date data type. Simplified version of `java.util.Date`.

```
public class Date implements Comparable<Date>
{
   private final int month, day, year;

   public Date(int m, int d, int y)
   {
      month = m;
      day   = d;
      year  = y;
   }

   public int compareTo(Date that)
   {
      if (this.year  < that.year ) return -1;
      if (this.year  > that.year ) return +1;
      if (this.month < that.month) return -1;
      if (this.month > that.month) return +1;
      if (this.day   < that.day  ) return -1;
      if (this.day   > that.day  ) return +1;
      return 0;
   }
}
```

only compare dates to other dates

Using Comparable

- Check if an array is in ascending order:
    - Ascending order: {1, 5, 10, 20}, etc.
    - Not in ascending order: {3, 2, 1}, {1, 10, 2}, etc.

```
public class Ascending {
    public static boolean ascending(Comparable[] a) {
        for (int i = 0; i < a.length - 1; i++) {
            if (a[i].compareTo(a[i + 1]) > 0) return false;
        }
        return true;
    }

    public static void main(String[] args) {
        Integer[] a = { 1, 2, 3, 3 };
        StdOut.println(ascending(a));
    }
}
```

Equals

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`.

equivalence relation

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.

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)
   {
      if (this.day   != that.day  ) return false;
      if (this.month != that.month) return false;
      if (this.year  != that.year ) return false;
      return true;
   }
}
```

check that all significant fields are the same

Implementing equals for user-defined types

Seems easy, but requires some care.

typically unsafe to use equals() with inheritance (would violate symmetry)

```
public final class Date implements Comparable<Date>
{
   private final int month;
   private final int day;
   private final int year;
   ...

   public boolean equals(Object y)
   {
      if (y == this) return true;

      if (y == null) return false;

      if (y.getClass() != this.getClass())
         return false;

      Date that = (Date) y;
      if (this.day   != that.day  ) return false;
      if (this.month != that.month) return false;
      if (this.year  != that.year ) return false;
      return true;
   }
}
```

must be Object.
Why? Experts still debate.

optimize for true object equality

check for null

objects must be in the same class
(aside: getClass() vs. instanceof)

cast is guaranteed to succeed

check that all significant fields are the same

Equals design

"Standard" recipe for user-defined types.

- Optimization for reference equality.
- Check against `null`.
- Check that two objects are of the same type and 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 (or otherwise deal with -0.0 and NaN)

← apply rule recursively

← can use `Arrays.deepEquals(a, b)` but not `a.equals(b)`

Best practices.

- No need to use calculated fields that depend on other fields.
- Compare fields mostly likely to differ first.
- Make `compareTo()` consistent with `equals()`.

e.g., cached Manhattan distance

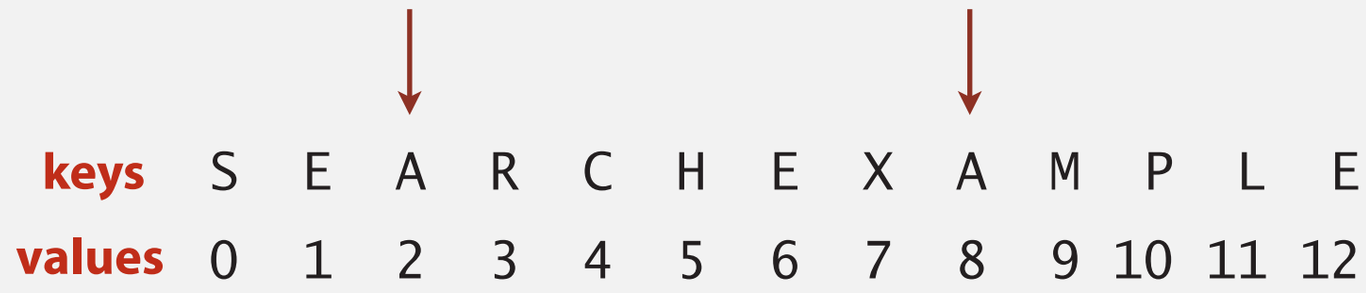
`x.equals(y)` if and only if `(x.compareTo(y) == 0)`

Back to Symbol Tables!

ST test client for traces

Build ST by associating value i with i^{th} string from standard input.

```
public static void main(String[] args)
{
   ST<String, Integer> st = new ST<String, Integer>();
   for (int i = 0; !StdIn.isEmpty(); i++)
   {
     String key = StdIn.readString();
     st.put(key, i);
   }
   for (String s : st.keys())
      StdOut.println(s + " " + st.get(s));
}
```



output

```
A  8
C  4
E  12
H  5
L  11
M  9
P  10
R  3
S  0
X  7
```

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

% java FrequencyCounter 1 < tinyTale.txt
it 10

% java FrequencyCounter 8 < tale.txt
business 122

% java FrequencyCounter 10 < leipzig1M.txt
government 24763
```

minimum length

tiny example
(60 words, 20 distinct)

real example
(135,635 words, 10,769 distinct)

real example
(21,191,455 words, 534,580 distinct)

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>();
      while (!StdIn.isEmpty())
      {
         String word = StdIn.readString();
         if (word.length() < minlen) continue;
         if (!st.contains(word)) st.put(word, 1);
         else                    st.put(word, st.get(word) + 1);
      }
      String max = ""; // Tracking key with maximum occurrences
      st.put(max, 0);  // So all other keys are greater
      for (String word : st.keys())
         if (st.get(word) > st.get(max))
            max = word;
      StdOut.println(max + " " + st.get(max));
   }
}
```

create ST

ignore short strings

read string and update frequency

print a string with max freq

http://algs4.cs.princeton.edu

3.1 Symbol Tables

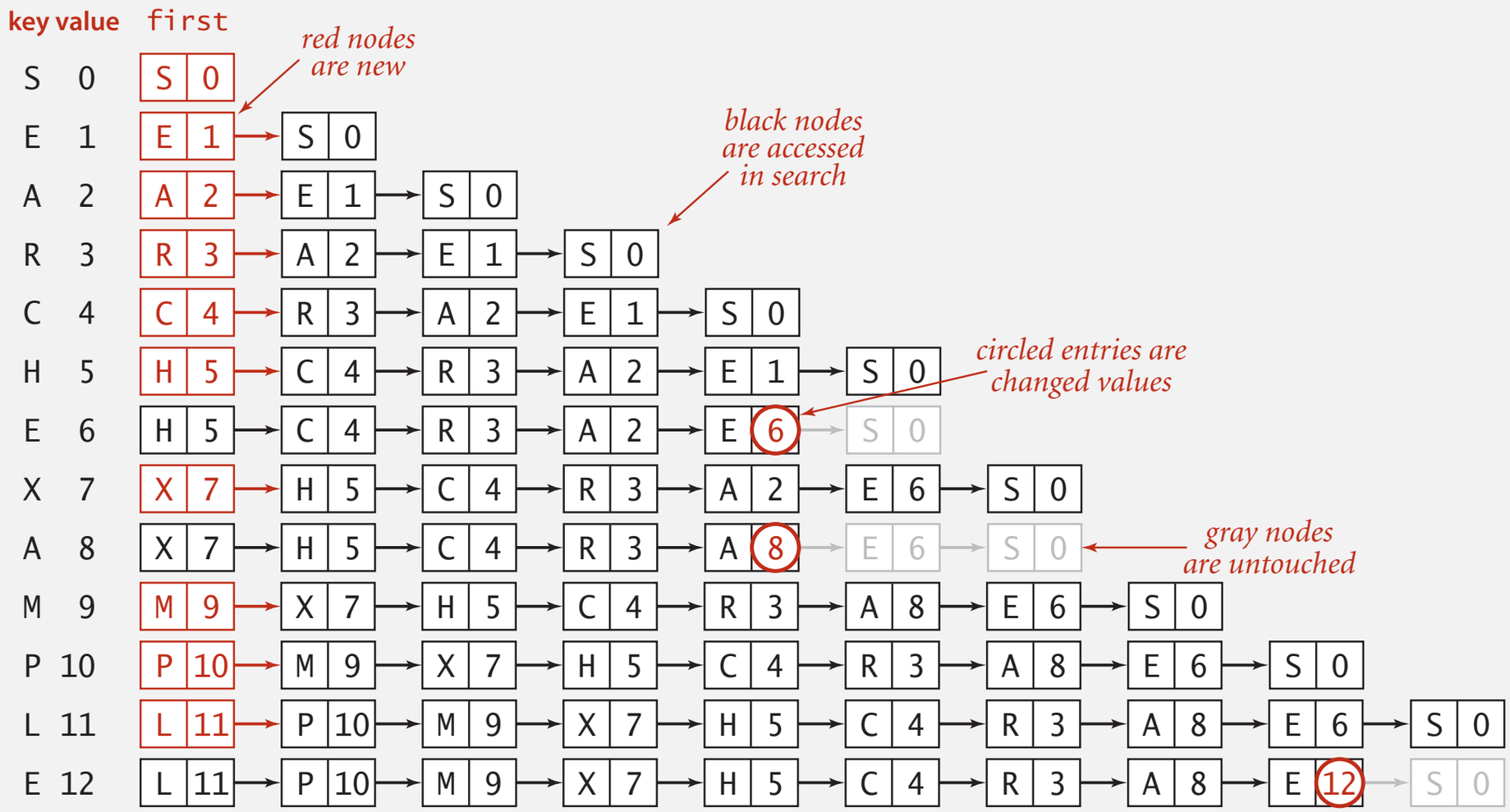
- *API*
- *elementary implementations*
- *ordered operations*

Sequential search in a linked list

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.



Trace of linked-list ST implementation for standard indexing client

Elementary ST implementations: summary

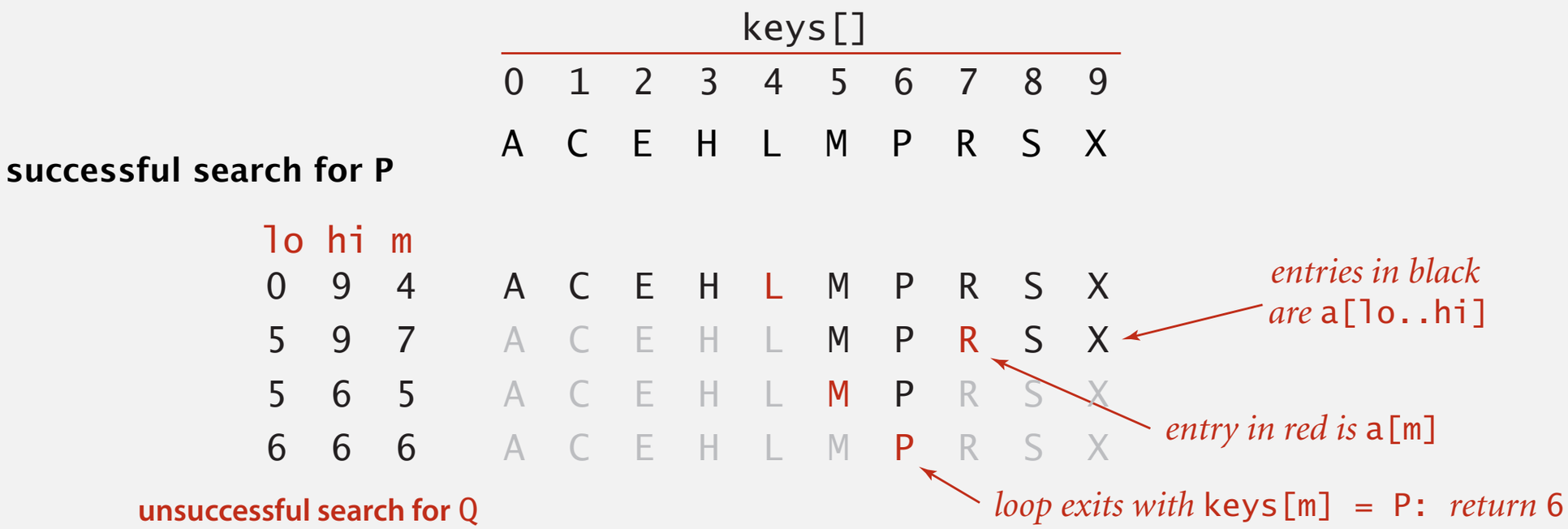
ST implementation	guarantee		average case		key interface
	search	insert	search hit	insert	
sequential search (unordered list)	N	N	$N / 2$	N	`equals()`

Challenge. Efficient implementations of both search and insert.

Binary search in an ordered array

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

Rank helper function. How many keys < k?



unsuccessful search for Q

Binary search: Java implementation

```
public Value get(Key key)
{
   if (isEmpty()) return null;
   int i = rank(key);
   if (i < N && keys[i].compareTo(key) == 0) return vals[i];
   else return null;
}
```

```
private int rank(Key key)                          number of keys < 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 mid;
   }
   return lo;
}
```

Binary search: trace of standard indexing client

Problem. To insert, need to shift all greater keys over.

key	value
S	0
E	1
A	2
R	3
C	4
H	5
E	6
X	7
A	8
M	9
P	10
L	11
E	12

keys[]									
0	1	2	3	4	5	6	7	8	9
S									
E	S								
A	E	S							
A	E	R	S						
A	C	E	R	S					
A	C	E	H	R	S				
A	C	E	H	R	S				
A	C	E	H	R	S	X			
A	C	E	H	M	R	S	X		
A	C	E	H	M	P	R	S	X	
A	C	E	H	L	M	P	R	S	X
A	C	E	H	L	M	P	R	S	X

entries in red were inserted

entries in gray did not move

N	vals[]									
	0	1	2	3	4	5	6	7	8	9
1	0									
2	1	0								
3	2	1	0							
4	2	1	3	0						
5	2	4	1	3	0					
6	2	4	1	5	3	0				
6	2	4	(6)	5	3	0				
7	2	4	6	5	3	0	7			
7	(8)	4	6	5	3	0	7			
8	8	4	6	5	9	3	0	7		
9	8	4	6	5	9	10	3	0	7	
10	8	4	6	5	11	9	10	3	0	7
10	8	4	(12)	5	11	9	10	3	0	7
	8	4	12	5	11	9	10	3	0	7

entries in black moved to the right

circled entries are changed values

Elementary ST implementations: summary

ST implementation	guarantee		average case		key interface
	search	insert	search hit	insert	
sequential search (unordered list)	N	N	$N / 2$	N	`equals()`
binary search (ordered array)	$\log N$	N	$\log N$	$N / 2$	`compareTo()`
In a future lecture...	$\sim \log N$	$\sim \log N$	$\sim \log N$	$\sim \log N$	`compareTo()`

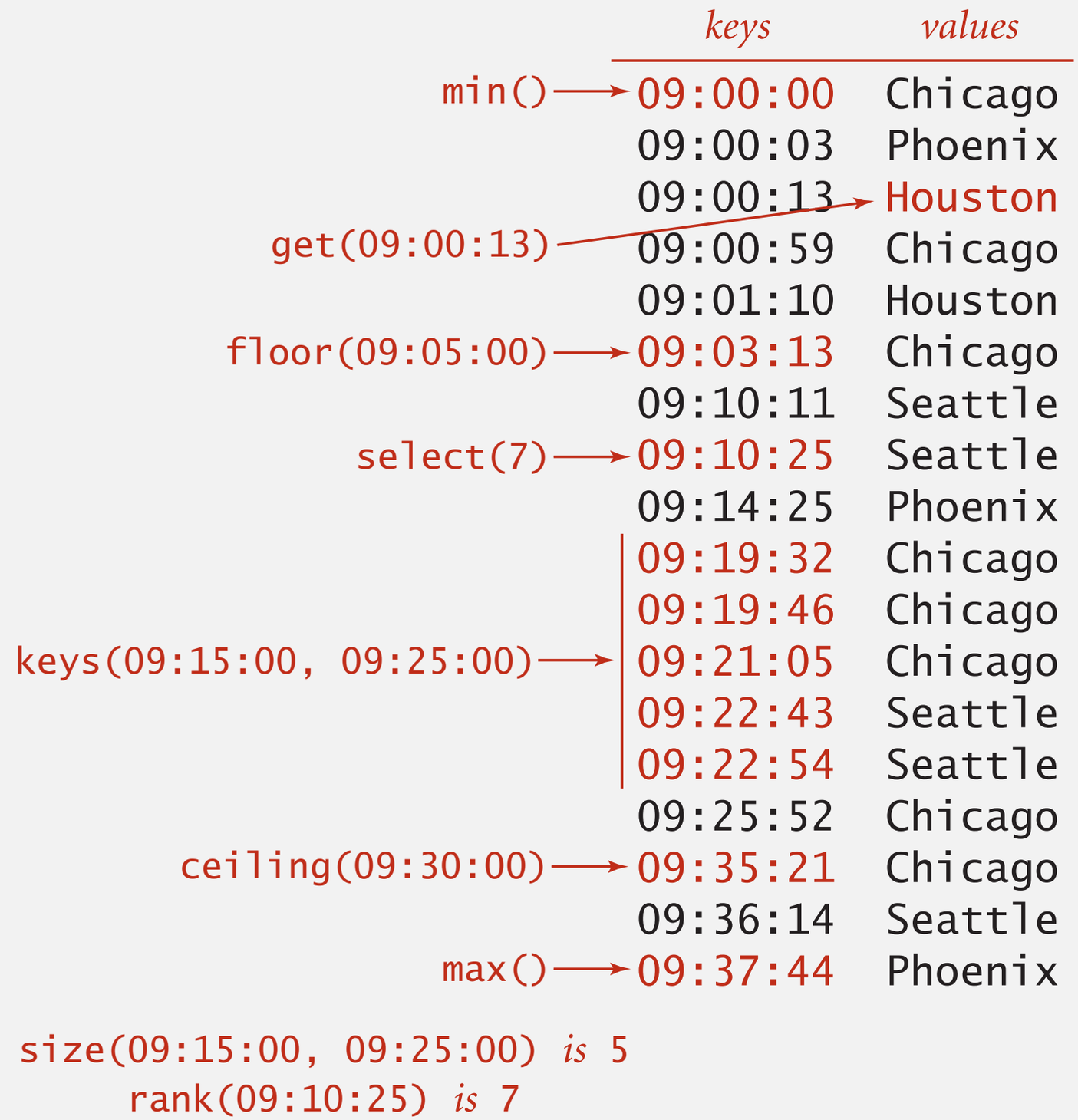
Challenge. Efficient implementations of both search and insert.

http://algs4.cs.princeton.edu

3.1 Symbol Tables

- API
- elementary implementations
- ordered operations

Examples of ordered symbol table API



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*
void	deleteMin()	*delete smallest key*
void	deleteMax()	*delete largest key*
int	size(Key lo, Key hi)	*number of keys between lo and hi*
Iterable<Key>	keys()	*all keys, in sorted order*
Iterable<Key>	keys(Key lo, Key hi)	*keys between lo and hi, in sorted order*

Symbol table implementations: summary

	unordered list implementation (sequential search)	ordered array implementation (binary search)
search	N	$\log N$
insert / delete	N	N
min / max	N	1
floor / ceiling	N	$\log N$
rank	N	$\log N$
select	N	1
ordered iteration	$N \log N$	N

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