

COMP 161 - Lecture Notes 14 - Search and Sort

In these notes we apply the basic recursive and iterative design template to the problems of searching and sorting.

Structural Recursion and `std::vector`

Structural recursions is about making recursive procedures that recurse on the recursive structure of the data. The most basic form of this structure come from having an empty collection `BASE CASE` and a non-base case where the data is deconstructed into the first element and all of the rest. The “rest” is a smaller version of the original structure. More generally, we can identify any non-recursive smallest size¹ as a base case and deconstruct the non-base case in any way so long as repetition of the decomposition eventually results in a base case². Such structures exists abstractly for just about any collection type. However, not all collections support recursive decomposition or do so in an inefficient manner. The C++ `std::vector` does not support recursive decomposition directly.

Thankfully, any structure with indexed elements can be managed recursively thought the recursive handling of the set of index values. A vector of size s used the integer interval $[0, s)$ for its indices. When the vector is empty, then the interval $[0, 0)$ is also empty³. The first of this interval is 0 and the rest is $[1, s)$. Similarly, the last is $s - 1$ and all but the last is $[0, s - 1)$. We can generalize this for any range of contiguous positive integers $[a, b)$. The interval is empty if $a \geq b$. When $a < b$, the first is a and the rest is $[a + 1, b)$.

To recursively process a vector we must write a procedure that takes the index interval bounds *first* and *last*. We don't always need both bounds. Pure functions on vectors can often exclude the last when recursing first to last or last when recursing last to first. Having both, however, provides maximum flexibility. If you need to mutate the vector and doing so changes the size, then you're probably going to need to shift first and last to account for the change in the vector's structure. You also can design with both bounds so that you can defer choosing the exact pattern of recursion to implementation time. Sometimes you don't realize that a pattern won't work until you really start working with the details.

If your goal is to work with a vector in its entirety, then the extra parameters change the basic interface to that problem. To hide them we use a `OVERLOADED` function where one version takes on the vector and calls the recursive procedure with first equal to zero and last equal to the size of the vector. This will cause the recursive variant of the function to consider all the elements of the vector.

¹ one element, two elements, etc

² all but the last + last, left half + right half, etc.

³ $[a, b)$ is empty if $a \geq b$

Figure 1 gives the basic template for first + rest structural recursion on vectors with a top-level variant that works on the whole vector by using the recursive variant as a helper. We'll be applying this template to search and sort.

```

1  /**
2   *
3   * @param v the vector
4   * @param first the smallest index to be processed
5   * @param last the excluded upper bound of the interval of indexes
6   *           to be
7   *           processed
8   * @return ...
9   * @pre 0 <= first, last < v.size()
10  * @post ...
11  */
12
13  ... foo([const] std::vector< ... >& v, int first, int last){
14
15      if( first >= last ){
16          // base case
17          ....
18      }
19      else{
20          ... v[first] ... foo(v,first+1,last)
21      }
22  }
23
24  /**
25   *
26   * @param v the vector
27   * @return
28   * @post ...
29   */
30  ... foo([const] std::vector< ... >& v){
31      ... foo(v,0,v.size()) ...
32  }

```

Figure 1: Structural Recursion Template for Vectors

Search: The Problem

Search is a fundamental problem in computing. Given a collection⁴, find a specific element, typically called the *search key*⁵ in that collection. Several variations can occur: find the first occurrence, the last occurrence, and more generally the k – *th* occurrence. With the *std::vector*, we want to return the index of occurrence. A simpler version of search is the *contains* predicate that returns true if the col-

⁴ *std::vector* for now

⁵ or just *key*

lection contains at least one occurrence of the search key. Finally, another search-like procedure is a counting procedure that returns the total number of occurrences.

For these notes we'll look at the version of search that examines the entire vector and returns the index of the first occurrence of the key. The declaration for this function is given in figure 2 with tests in figure 3.

```

1  /**
2   * Compute the location of the first occurrence of the integer key
3   * in the vector data.
4   * @param data vector of integers
5   * @param key search value
6   * @return -1 if key is not found, otherwise the index where key
7   * is first found
8   * @pre none
9   * @post none
10 */
11 int search(const std::vector<int>& data, int key);

```

Figure 2: The Basic Search Function

```

1  TEST(search, all){
2
3   EXPECT_EQ(-1, search(std::vector<int>({}), 1));
4   EXPECT_EQ(-1, search(std::vector<int>({2}), 1));
5   EXPECT_EQ(0, search(std::vector<int>({1}), 1));
6   EXPECT_EQ(1, search(std::vector<int>({1,3,5}), 3));
7   EXPECT_EQ(0, search(std::vector<int>({1,3,1}), 1));
8   EXPECT_EQ(2, search(std::vector<int>({1,3,5}), 5));
9
10 }

```

Figure 3: Tests for Basic Search

This basic interface and the tests should work regardless of the underlying implementation technique. How you solve the problem does not change the problem itself. In this case, we're interested in searching an entire vector and this function captures that problem succinctly.

Search: The Solutions

The recursive and iterative versions share a lot in common. We'll start with the recursive variant and then look at the iterative version. In both cases it's clear that we only need read-only access to the vector and using *pass-by-const-reference* would be a good idea.

Recursive Search

For the recursive implementation we need the variant of the search that accepts the bounds of the search range in order to recurse along that interval as shown in figure 4. It's worth noting this function, as declared, doesn't need to be recursive. We can, and should, view it as a more general version of search: find the key within this subsection of the vector. The more general problem of searching a part of a vector gives us the flexibility we need to enable recursion.

```

1  /**
2   * Compute the location of the first occurrence of the integer key
3   * in the vector data for the index range [fst,lst).
4   * @param data vector of integers
5   * @param fst the lower bound of the search range
6   * @param lst the excluded upper bound of the search range
7   * @param key search value
8   * @return -1 if key is not found, otherwise the index where key
9   * is first found
10  * @pre fst <= lst
11  * @post none
12  */
13 int search(const std::vector<int>& data, int fst, int lst, int key);

```

Figure 4: Recursive-Capable Search

When testing the more generic search we should test it not only for whole vector searches, the problem we originally set out to solve, but for partial vector searches as we see in figure 5. Test the procedure as it stands on its own, not just for some subset of its usage.

```

1  TEST(search, some){
2
3   // search all
4   EXPECT_EQ(-1, ln14::search(std::vector<int>({}), 0, 0, 1));
5   EXPECT_EQ(1, ln14::search(std::vector<int>({1, 2, 3, 2, 1}), 0, 5, 2));
6   EXPECT_EQ(-1, ln14::search(std::vector<int>({1, 2, 3, 2, 1}), 0, 5, 7));
7   // search some
8   EXPECT_EQ(3, ln14::search(std::vector<int>({1, 2, 3, 2, 1}), 2, 5, 2));
9   EXPECT_EQ(-1, ln14::search(std::vector<int>({1, 2, 3, 2, 1}), 2, 3, 2));
10
11 }

```

Figure 5: Recursive-Capable Search Tests

The top-level search⁶ simply calls the more generic version with the interval for the entire vector as shown in figure 6.

To work out the recursive implementation of the generalized search we start with the base case. When a vector is empty, it can-

⁶ search the whole vector

```

1 int search(const std::vector<int>& data, int key){
2     return search(data, 0, data.size(), key);
3 }

```

Figure 6: Top-Level Search: Recursive Implementation

not contain the key so return -1 . With the non-empty case we work out the problem in terms of the first⁷ and the result of recursing on the rest⁸. In the context of search this boils down to the following observation: either the first is the key, the key is in the rest, or the key is not in the vector at all. A complete case analysis reveals four possible situations⁹: the first is the key and the rest contains the key, the first isn't the key and the rest contains the key, the first is the key and the rest doesn't contain the key, or the key is neither the first nor is it in the rest. By recursively calling search on the rest we expect to get the location of key in the rest or a -1 if it's not in the rest. For example, if the vector v contains $\{2, 3, 2, 1\}$ and the search key is 2 then the recursive call $search(v, 1, v.size(), 2)$ should return 2, but because 2 is also at location 0, the first, we should return 0 in favor of 2. If we were searching for the key 4 then the recursive call should return -1 and given that the first, 2 isn't 4, we should return -1 . By continuing to work examples like this we see that the recursion is necessary if and only if the first isn't the key. In the case where the first element is the search key, we should just return the first index without recursively searching the rest.

⁷ data[fst]⁸ search(data, fst+1, lst, key)⁹ two potential locations (first and rest) with two potential states (contains and doesn't contain)

In figure 7 we see the finished code for the recursive search. The two cases of the non-empty portion of the main conditional have been flattened into the main conditional itself rather than nesting a second conditional in the else of the main conditional.

```

1 int search(const std::vector<int>& data, int fst, int lst, int key){
2     if( fst >= lst ){
3         return -1;
4     }
5     else if( data[fst] == key ){
6         return fst;
7     }
8     else{
9         return search(data, fst+1, lst, key);
10    }
11 }

```

Figure 7: Search Some: Recursive Implementation

It's worth stopping to check our implementation against our case analysis. The most important case is the recursive one: the search

range of the vector is either empty or it's not. Empty case is handled by the *if* and the non-empty case is caught either by the *else if* or the *else*. When the search range isn't empty we can condense the four cases down to two: the key is the first or its existence and location can be determined through recursively searching the rest. The *else if* case should catch the case where the first is the key. The *else* covers the other case.

Iterative Search

The basic logic of iteration is to traverse the structure and accumulate the solution while you traverse. Vector traversals work the same way as string traversals. You simply count your way through the set of index values¹⁰. By default we start at zero and count up, but other counting schemes are possible. The solution to this problem is an index value or -1. Developing the recursive solution showed us that we shouldn't need to continue searching once we discover the first occurrence of the search key. In an iterative world this means stopping the traversal and returning the current index. Combine this with the fact that we need to accumulate an index value, we notice that an extra accumulator isn't required, the loop is already accumulating what we need in its counter. We need to be careful though. Basic accumulator logic dictates that we return the accumulated value when traversal is done. When we don't find the search key the loop's counter will be the size of the vector and we want -1. Similarly, if the vector is empty, the initial counter value is 0 and we need to return -1. The fix here is to recognize that when we find the key, we can return the counter, and if we don't find the key we just return the literal -1.

In figure 8 we see the finished code for the recursive search. The two cases of the non-empty portion of the main conditional have been flattened into the main conditional itself rather than nesting a second conditional in the else of the main conditional.

¹⁰ [0,size)

```

1 int search(const std::vector<int>& data, int fst, int lst, int key){
2
3     for(unsigned int i{0}; i < data.size() ; ++i){
4         if( data[i] == key ){
5             return i;
6         }
7     }
8
9     return -1;
10 }
```

Figure 8: Search: Iterative Implementation

It's worth your time to really analyze this solution and clearly identify how we tweaked the basic "traverse and accumulate" logic to arrive at this implementation. Imagine we're searching the vector v containing $\{2, 3, 1, 2\}$ for the search key 1. If we're currently looking at index $i = 1$ then we've accumulated the fact that everything before 1 doesn't contain the key¹¹. The vector element at 1 isn't the key, so we should leave the accumulated value as -1 . If the key were 2 then we should have discovered the key on a previous iteration and the accumulator would be 0, the location of the first occurrence of the key in the previously traversed structure. Finally, consider the case where the key is 2 and the current location is 3. We should have discovered 2 at location 0 on a previous iteration and upon rediscovering 2 and location 3 we should preserve the previous accumulator value because our goal is to find the first occurrence. If you wanted to capture this kind of on the nose iterative thinking in code you'd end up with what you see in figure 9. From this perspective our version is an optimization of the full traversal version¹².

¹¹ i.e. our accumulator is -1

¹² why keep going once you've found it and why accumulate the location twice, once for the traversal loop and once for the iteration
Figure 9: Search: Iterative Implementation with explicit accumulation and full traversal

```

1 int search(const std::vector<int>& data, int fst, int lst, int key){
2
3     int loc{-1};
4     for(unsigned int i{0}; i < data.size() ; ++i){
5         if( data[i] == key && loc == -1 ){
6             loc = i;
7         }
8     }
9
10    return loc;
11 }
```

Sort: The Problem

Sorting is something we teach small children before they even learn to read and do arithmetic. The problem is that fundamental to our thinking. It may come as a surprise then that it provides a wide array of variations. For starters, we can sort in ascending or descending order. As a computing problem we can think of it as a function and a mutator. The the case of the former, we'd be producing a second vector with the same contents as the argument but a potentially different order. As a mutator we would be committing to rearranging the contents of an existing vector. Of course, we could implement a function via mutation of a local vector and we can implement a mutator by making a local sorted copy then swapping the original for

the contents of the sorted copy.

The version of sort we'll be looking at is an ascending order, IN-PLACE sort. Working in-place is just another way of saying we want to implement sort as a vector mutator and manage the mutation directly rather than copying the contents and modifying the copy. In-place sorts are important to computing because they guarantee we do not make copies of data and in general, copying can be costly.

Another desirable property of sorts is STABILITY. A stable sort will leave equivalent values in the same relative order as they were originally¹³. We'll set aside stability as a design goal for now but should step back from whatever implementation we end up with and determine if we just so happened to produce a stable sort.

¹³ the first 1 is still first, the second 1 is still second, and so on

In figure 10 we see the declaration for a sort mutator. The accompanying tests are given in figure 11. When developing tests it's good to work examples in increasing order of complexity. This typically boils down to increasing sizes and maybe different cases for a particular size. For vectors of size zero or one, the sort seems trivial. There's really nothing to do. For a size of two we can get a feel for when work does and does not need to happen because sometime we need to swap the order and sometimes we don't. Finally, vectors of size greater than run a wide gambit of nearly sorted to totally unsorted.

Figure 10: Sort as a Mutator

```

1 /**
2  * Sort the contents of data in least to greatest order
3  * @param data vector of integers
4  * @return none
5  * @pre none
6  * @post contents of data have been sorted in least to greatest
   order
7  * for all i in [0,data.size()-1], data[i] <= data[i+1]
8  */
9 void sort(std::vector<int>& data);

```

This time we'll start by looking at the iterative implementation because loops and iteration play nice with mutation and perhaps we'll have an easier time teasing out basic logic in that environment. Once we've teased out some iterative sorting logic, we'll see if that doesn't provide insight into a recursive strategy.

Iterative Implementation

Iterative procedures still have base cases, cases where no iteration¹⁴ is needed. The question we ask ourselves is: when can we tell that

¹⁴ traverse and accumulate

```
1 TEST(sort,all){
2
3     std::vector<int> sortme;
4
5     sort(sortme);
6     EXPECT_EQ(std::vector<int>({}),
7         sortme);
8
9     sortme = std::vector<int>({1});
10    sort(sortme);
11    EXPECT_EQ(std::vector<int>({1}),
12        sortme);
13
14    sortme = std::vector<int>({7,4});
15    sort(sortme);
16    EXPECT_EQ(std::vector<int>({4,7}),
17        sortme);
18
19    std::vector<int> data{8,7,6,5,4,3,2,1};
20    sort(data);
21    EXPECT_EQ(std::vector<int>({1,2,3,4,5,6,7,8}),
22        data);
23
24 }
```

Figure 11: Sort as a Mutator

a vector is sorted without actually traversing the vector in any way? The answer is when the vector contains one or fewer items. Here we see an instance of a non-empty base case: Vectors of size one need not be worked on because they are trivially sorted already. As we move to implementing our sort we want to ensure that our traversal loop only traverses vectors with size greater than one.

Iterative mutation is about accumulating effect, not value. If we've done i steps of the sorting traversal then the vector should be sorted in the first i places. This also means that when the loop is done, the sort should be done. There's nothing left to do. We need to combine this mutation based thinking with the base case now. If the base case includes regions of size one, then our traversal should begin at the second element in the vector and accumulate effect across the whole region of size two. This all pushes us to a loop design that begins at $i = 1$ and counts up towards the size of the vector. Beginning at one means the loop will only perform iterations for vectors with a size greater than one¹⁵.

The last thing we need to work out is how to accumulate. Let's look an example. If the vector v originally contained $\{3, 7, 1, 2, 8, 6\}$ and we've performed 2 iterations, then the current index i should be 3 and v from 0 to 2 should contain $\{1, 3, 7\}$. We're now faced with the problem of modifying the region from 0 to 3 such that the 2 at location 3 is placed at location 1 and the 3 and 7 following it are shifted over to the right one. It is not immediately clear how to do this.

Because we're trying to work in-place, it's important that we think in terms of moving data within the existing vector and not using any kind of operation that changes the size of the vector. This rules out the use of the `std::vector` *insert* and *erase* methods. There is an *assign* method¹⁶ that might be useful, but in order to use it we have to know where, exactly, things need to go. When we look at our example, we see what needs to go where because our visual processing lets us work out the global structures. The computer doesn't see that. It sees nothing. The computer has a sorted region and an element right next to that region that needs to get moved so that the whole region is now sorted. There appears to be no pre-build solution to this¹⁷, so the answer is, of course, to abstract away the problem in the form of a helper.

When faced with a task for which no known method, procedure, or operation provides an solution you should quickly design a helper to solve the problem. This is the bread and butter of top-down design¹⁸. Find the helpers you need by working them out in context. What we need is a procedure that takes the vector and the bounds of a region of that vector. To shake things up we'll include the up-

¹⁵ assuming we continue as long as $i < \text{size}$ of course

¹⁶ <http://www.cplusplus.com/reference/vector/vector/assign/>

¹⁷ or we just want to solve this problem on our own anyway

¹⁸ The real win here is that you're able to solve the problem without actually solving the problem by clearly identifying, naming, and specifying the missing piece(s)

per bound of the region this time. So, our precondition is that for the region contained in $[fst, lst]$, the data in $[fst, lst)$ is sorted. We know nothing about the element at $lst - 1$. The procedure should then move the data at $lst - 1$ such that everything in $[fst, lst)$ is now sorted. We're also assuming that the sorted region contains at least one item. This means $fst < lst$. In short, this procedure inserts the item at lst into the region to the right of lst and that that ends at fst so that the whole region is sorted. We can now properly declare and document this as a C++ procedure as shown in figure 12.

```

1  /**
2   * Move data[lst] into sorted region data[fst..lst-1] such that
3   * the whole region is sorted.
4   * @param data vector of integers
5   * @param fst lowest index of region for insertion
6   * @param lst the location of the item to be inserted. Also
7   * one more than the last of the insertion region
8   * @pre fst < lst. data[fst .. lst-1] is sorted in
9   * least to greatest order
10  * @post data[fst..lst] is sorted in least to greatest order
11  */
12 void insert(std::vector<int>& data,
13            unsigned int fst, unsigned int lst);

```

Figure 12: Insert: Declaration

A few tests for insert are given in figure 13. Once again, develop tests ranging from simple base cases to larger more complex cases. We should also test insert in a more general setting. We expect the value of fst to always be 0 but it doesn't need to be. Testing more general cases lets you re-evaluate and reaffirm the pre and post conditions.

Now that it's clear what *insert* should do and how to use it, we can finish up *sort*. In figure 14 we see the complete iterative sort known as INSERTION SORT. Insertion sort is a class, well known algorithm for sorting algorithm. Our design uses an insert helper. It's not uncommon to see the helper replaced by the actual insertion logic. This design clearly separates the core iterative logic of the sort from the accumulative operation carried out by insert. We're also able to express *sort* without knowing how to insert. In short, this design clearly separates different concerns of the problem when compared to the explicit, all-in-one version you'll find in texts and online.

We now need to finish the implementation of *insert* if we want sort to actual work. If our sort is to be an in-place mutator, then *insert* must also work in-place through mutation. Once again, we'll choose

```

1 TEST(insert,all){
2
3     std::vector<int> testme({3,2});
4     ln14::iter::insert(testme,0,1);
5     EXPECT_EQ(std::vector<int>({2,3}),
6               testme);
7
8     testme = std::vector<int>({2,3});
9     ln14::iter::insert(testme,0,1);
10    EXPECT_EQ(std::vector<int>({2,3}),
11              testme);
12
13    testme = std::vector<int>({2,4,5,7,2,4});
14    ln14::iter::insert(testme,0,4);
15    EXPECT_EQ(std::vector<int>({2,2,4,5,7,4}),testme);
16
17    testme = std::vector<int>({9,4,5,7,2,1});
18    ln14::iter::insert(testme,1,4);
19    EXPECT_EQ(std::vector<int>({9,2,4,5,7,1}),testme);
20
21 }

```

Figure 13: Insert: Tests (Iterative Version)

```

1 void iter::sort(std::vector<int>& data){
2
3     for(unsigned int i{1}; i < data.size(); i++){
4         iter::insert(data,0,i);
5     }
6     return;
7 }

```

Figure 14: Sort: Iterative Sorting

to use iteration rather than recursion in order to implement insert.

The first thing we need to recognize is that *insert* needs to traverse and accumulate through the portion of the vector indexed by $[fst, lst)$. The in-place requirement really pushes us away from the standard fst to lst traversal though. Let's see why. Say some vector v contains $\{1, 3, 5, 7, 2\}$ in the $[fst, lst]$ range¹⁹. After a single iteration from fst to lst , what should the vector range look like? What we need to remember is that the region we've traversed should be the final solution with respect to that portion of the vector. So after a single iteration the fst should still be 1. After two iterations we should see $\{1, 2, 3, 5, 7\}$ because the 2 needs to be at $fst + 1$, the second spot, in the final solution. If we're to do this in place, then we need to shift everything in $[i, lst)$ to the right one so that $v[lst]$ can go at location i . Once again, it seems like we'd need a helper for this because there's no operation for shifting a region. Before we go down that path, let's explore a different traversal pattern: $lst-1$ to fst .

¹⁹ So $v[fst]$ is 1 and $v[lst]$ is 2

When traversing starting at $lst-1$ and working down towards fst we need not only rethink the loop counting pattern but the iterative logic. The region we've traversed already is not at the back of the range, not the front. After one iteration with our same example vector v , what should v look like? The basic idea now is that if we're dealing with location i on the current iteration that after this iteration $[i, lst]$ should be sorted and contain all the values that were originally in $[i, lst]$. So after one iteration our vector should look like $\{1, 3, 5, 2, 7\}$. What we've actually done is reversed the problem a bit. We want to insert the value at lst into the sorted region $[fst, lst)$. In working lst to fst , we actually inserting the last of $[fst, i]$ into the sorted region $(i, lst]$. Initially the region $(i, lst]$ is just lst . After the first iteration, it's $[lst - 1, lst]$, and then on down.

The fact that both regions are sorted lets us cut some corners. Let's trace insert a bit to find it. After one iteration our example region of v contains $\{1, 3, 5, 2, 7\}$. We can accomplish this by swapping the last two elements or more generally $v[i]$ and $v[i + 1]$. On the second iteration we need to "insert 5 into $\{2, 7\}$ ". Once again a simple swap would accomplish this. In fact, if you continue on you'll find that all we ever need to do is swap $v[i]$ with $v[i + 1]$ until the original insertion number, 2, gets to its final location at which point we can just stop traversing all together. Why is this? Until we get the original $v[lst]$ to the correct spot, $v[i + 1]$ is always that number and everything to its right (if it exists) must be greater than or equal to all the stuff in $[fst, i]$ because that makes up the original sorted region $[fst, lst)$! When we swap $v[i]$ with $v[i + 1]$ we're simply putting a number back where it was but shift to the right of the original $v[lst]$ rather than the left of $v[lst]$ where it started. When $v[i]$ is less than

$v[i + 1]$, then we can stop the traversal because we know everything to the left of $v[i]$ (if it exists) is less than or equal to $v[i]$ and must therefore be less than $v[i + 1]$. By traversing from right to left rather than left to right reversed the problem a bit but in doing so naturally solved our shift and place problem by turning the shift into iterated swaps.

The final version of iterative *insert* is shown in figure 15. It's design and implementation merit repeated study as it illustrates how varying the traversal pattern can sometimes avoid some problems encountered with the standard left to right traversal at the cost of forcing us to reverse our thinking on the iterative logic of the problem. The other thing you'll see is careful management of counting down through an index range using unsigned integers. If *fst* is 0^{20} , then when the unsigned int *i* is equal to *fst* and we do *i* -- we won't get -1 but a very large number instead. The check against the size of the vector ensures we never carry out an iteration with an *i* value outside of $[fst, lst]$.

²⁰ and it always will be for our target usage of insert

```

1 void iter::insert(std::vector<int>& data,
2     unsigned int fst, unsigned int lst){
3
4     for(unsigned int i{lst-1}; i >= fst && i < data.size(); i--){
5         if( data[i+1] < data[i] ){
6             std::swap(data[i],data[i+1]);
7         }
8         else{
9             return;
10        }
11    }
12    return;
13 }
```

Figure 15: Insert: Iterative Implementation

Recursive Implementation