CSE 331: Software Design & Implementation

Homework 5 (due Wednesday, November 1st at 11:00 PM)

The problems in this assignment contain a mix of written and coding parts. When you have completed them, submit your solutions in Gradescope. The written problems should be clearly labeled and submitted as a pdf to the "HW5 Written" assignment. The following completed files should be directly submitted for the coding portion to the "HW5 Coding" assignment:

color_list.ts	color_tree_test.ts	list_test.ts	index.tsx
color_list_test.ts	ui.tsx	parser.ts	
color_tree.ts	ui_test.tsx	parser_test.ts	

For complete instructions for how to submit assignments in this course see the Homework Turn-in Guide.

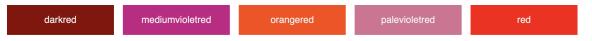
Reminder: Mutation is *not* allowed. Only write straight-line code, conditionals, and recursion. This applies to *all* code written for this assignment, including tests.

Start by checking out the starter code using the command

```
git clone https://gitlab.cs.washington.edu/cse331-23au-materials/hw-highlight.git
```

Then, install the modules using npm install --no-audit.

Try out the application by running npm run start and navigating to localhost:8080. The app contains a text area which allows the user to search for colors whose names contain a given sub-string. For example, searching for "Red" will show "DarkRed", "PaleVioletRed", etc.



The app contains another text area which allows users to draw text with different colored highlighting. The text area expects each line of input to start with the name of a color followed by the text to be highlighted in that color. For example, the input on the left produces the output on the right:

MediumPurple hello CornflowerBlue , LightSeaGreen world DarkOrange !



All of the color possibilities are listed in colors.ts. More details on the contents of this file and others are to follow, but browse the starter code briefly to see what we're working with.

1. List and Shout (20 points)

The following parts consist entirely of coding work. It should be submitted with "HW5 Coding".

The application described on the page prior records information about an individual color in a triple, called ColorInfo, of the form (n, c, w), where n is the name of the color, c is the CSS description of the color, and w is a boolean indicating whether text with that background color should be white rather than black. The full list of available colors is stored in a List<ColorInfo> called COLORS in colors.ts.

The file color_list.ts exports two functions, findMatchingNames and getColorCss, that search for information on colors. Currently, both operations work directly with the list COLORS. In this problem, we will replace this with an ADT so that we can easily change the representation used during searches later on.

- (a) Create a new interface ColorList in color_list.ts that has just the two operations mentioned above. Be sure that the interface and its functions have proper documentation.
- (b) Change the code in ui.tsx to import and use only the interface.

Specifically, change the two functions, showColors and showHighlights, to take a ColorList called "colors" as part of their props argument and then use that ColorList to perform the search operations. (You will need to change the helper functions of each of those functions as well.)

(c) Create a class in color_list.ts called SimpleColorList that implements the ColorList interface. Its constructor should take a List<ColorInfo> as an argument and store it in a field. The class should then implement the methods findingMatchingNames and getColorCss in the same way as the existing functions with those same names except that the methods will use their colors field rather than the COLORS constant. After adding these methods to the new class, you can remove the functions findingMatchingNames and getColorCss from the file. (The helper functions findMatchingNamesIn and getColorCssIn can stay as they are, however.)

Be sure to properly document the class and add an abstraction function of the class.

(d) Create a function, called makeSimpleColorList, that returns an instance of SimpleColorList that uses colors from the COLORS list. Be sure to write a complete specification for the function.

All instances of SimpleColorList created with the same argument (COLORS) will behave identically on all operations, so there is no need to create multiple instances. To avoid creating multiple instances, we can use the "singleton" pattern: define a constant that stores a single instance of the class and have makeSimpleColorList always return that instance.

Make sure that ColorList and makeSimpleColorList are the only symbols exported in the file. Clients should not be able to directly access the class, the singleton instance, or anything else.

(e) Change the tests in color_list_test.ts and ui_test.tsx to use makeSimpleColorList to create one instance of ColorList and use that instance to perform the operations throughout the tests.

Make sure that all the tests still pass by running npm run test. (Note that npm run start will still not work until we finish the next part.)

(f) Change the code in index.tsx to use makeSimpleColorList to create an instance of ColorList and pass that to the functions it calls from ui.tsx.

Make sure that the app still works by running npm run start and trying it out.

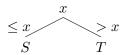
BSTs

The next problem makes use of the following inductive type, representing a binary search tree

type
$$BST := empty$$

| $node(x : \mathbb{Z}, S : BST, T : BST)$ with conditions A and B

where "A" is the condition that, for every $y : \mathbb{Z}$ with y > x, the value y cannot appear in any node of S, and "B" is the condition that, for every $y : \mathbb{Z}$ with $y \le x$, the value y cannot appear in any node of T. (We will define "appear in" more precisely below, but hopefully this is clear: numbers in the nodes of S must be less than or equal to x and numbers in the nodes of T must be greater than x.)



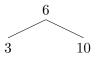
Conditions A and B are *invariants* of the BST node. Every node that is created must have these properties, and we are allowed to use the fact that they hold anywhere in our reasoning.

We can define the list of numbers appearing in the tree as follows:

 $\begin{aligned} & \textbf{func values(empty)} & := \text{ nil} \\ & \text{values(node}(x, S, T)) & := \text{ concat}(\text{values}(S), \text{cons}(x, \text{values}(T))) & \text{ for any } x : \mathbb{Z} \text{ and } S, T : \text{BST} \end{aligned}$

In words, the list of numbers in node(x, S, T) is the list of numbers appearing in S followed by x followed by the list of numbers appearing in T. (This is the natural way to order the list because it is then sorted.) For example:

B: BST = node(x: 6, S: node(x: 3, S: empty, T: empty), T: node(x: 10, S: empty, T: empty))



values(B) = cons(3, cons(6, cons(10, nil)))

Lists

We will also need the prefix and suffix functions from HW3, but we will combine them into a single function, called split, which is defined as follows:

In the third case, to split a non-empty list cons(a, L) into its first m + 1 elements and the rest, we recursively split L into its first m elements and the rest and then prepend a to the prefix, giving it length m + 1.

It is not hard to prove, using techniques we saw in HW3, that, if $m \leq \text{len}(L)$, then split(m, L) returns a pair (P, S) with the properties that L = concat(P, S), len(P) = m, and len(S) = len(L) - m. We will need those facts in the next problem. (You can cite this as "Lemma 1".)

2. Not My Cup of Tree (20 points)

The following parts consist entirely of written work. It should be submitted with "HW5 Written".

The function "values", defined on the previous page, turns a BST into a List, but we will also need a way to turn a List into a BST. Here is one way to do so:

This definition turns a non-empty list L into a tree by calling split(m, L) to split L into (P, S), each containing approximately half of L. Then, we further split S into its first element, b, and the rest, R. Finally, we return the tree with b at its root and P and R, each made recursively into trees, as its left and right subtree.

Note that, in order for node(b, makeBst(P), makeBst(R)) to satisfy the BST invariant, it must be the case that everything in P is less than or equal to b and everything in R is greater than b. That will always be the case if we start with a list that is **sorted**: the input list L = concat(P, cons(b, R)): is only sorted if everything in P is smaller than b and everything in R is larger than b.

We said "approximately half" above because, if len(L) is odd, then it cannot be split exactly in half. The definition above says to split after $m = \lfloor len(L)/2 \rfloor$ elements, where the notation $\lfloor \cdot \rfloor$ means to round *down* to the closest integer. In particular, when len(L) is odd, m will be (len(L) - 1)/2.

To prove that makeBst is correct, we want to show that it produces a tree containing the same elements as in the list it was given, i.e., that values(makeBst(U)) = U holds for any list U.

This is a little trickier than it initially appears, however, because makeBst(cons(a, L)) does not make a recursive call to makeBst(L). Instead, it makes recursive calls to makeBst(P) and makeBst(R), and all we know about the lists P and R is that they are shorter than cons(a, L). In order to apply the inductive hypothesis to P and R, we will need to phrase the claim a little differently.¹

We will prove by induction on $n : \mathbb{N}$ that values(makeBst(U)) = U holds for any list U with len(U) $\leq n$. This means the equation holds for any list U since we can apply the theorem with n = len(U), as we know that the condition len(U) $\leq n = \text{len}(U)$ certainly holds.

(a) The base case, P(0), says that values(makeBst(U)) = U for any list U with len(U) = 0. The only list satisfying this claim is nil (i.e., len(U) = 0 implies that U = nil).

Prove by calculation that values(makeBst(nil)) = nil.

(b) The inductive step, P(n + 1), says that values(makeBst(U)) = U for any list U with $len(U) \le n + 1$. Note that $len(U) = n + 1 \ge 1$, in which case, U = cons(a, L) for some some $a : \mathbb{Z}$ and L : List, and it is only necessary to prove the claim for a U of this form.

Prove by calculation that values(makeBst(cons(a, L))) = cons(a, L). Be sure to use the inductive hypothesis which tells us that values(makeBst(U)) = U for any list U with len $(U) \le n$.

That completes the induction argument. We have shown that values(makeBst(U)) = U holds for any list U.

(c) **Extra Credit**: The definition of makeBst only makes sense if S (the second half of the list) is non-empty. Otherwise, there would be no element "b" at the front of S to put at the root of the tree!

Use the properties of split mentioned on the previous page and the fact that $len(L) \ge 1$ to prove that $len(S) \ge 1$, where (P, S) = split(m, L) and m = |len(L)/2|. That means S has a first element.

¹This approach is sometimes called "strong" induction.

3. You Haven't Heard the Last of Tree (20 points)

The following parts consist entirely of coding work. It should be submitted with "HW5 Coding".

In an effort to make lookups more efficient, we decide to make a tree implementation of ColorList. Some portions have already been created or defined, so it's your job to put it all together.

(a) Translate the mathematical definition of makeBst from question 2 into a TypeScript function in color_tree.ts with the following signature:

```
export const makeBst = (L: List<ColorInfo>): ColorNode => {..};
```

ColorNode is a BST type for colors (rather than integers) provided in color_node.ts. node is a function to create a ColorNode provided in the same file. The function split is provided in list.ts. Note that colors are sorted by their string names, so there's no need to implement any additional ordering, you can use default string ordering with standard operations >, <, =.

Be sure to write a specification for the function.

(b) Write test cases for makeBst in the file color_tree_test.ts. There is one example test given. Follow the rules taught in lecture for choosing appropriate test cases.

Include brief comments in your test suite justifying the test cases you chose.

Confirm that all your tests pass by running npm run test.

Next, we consider the following function, which looks up a value in a binary tree:

func $lookup(y, empty)$:= undefined		for any $y:\mathbb{Z}$
lookup(y,node(x,S,T))	$:= \operatorname{node}(x, S, T)$	if x = y	for any $x, y : \mathbb{Z}$ and $S, T : BST$
lookup(y,node(x,S,T))	:= lookup(y,T)	if x < y	for any $x, y : \mathbb{Z}$ and $S, T : BST$
lookup(y,node(x,S,T))	:= lookup(y, S)	if y < x	for any $x, y : \mathbb{Z}$ and $S, T : BST$

(c) Translate the definition of lookup into a TypeScript function in the file color_tree.ts with the following signature:

export const lookup = (y: string, root: ColorNode): ColorInfo | undefined => {..};

Rather than implementing lookup with numbers, as given in the math definition, we will look up colors in a ColorNode BST. When we compare the value y, a color, to the current ColorNode, we just want to compare the color of the ColorNode's info. Then, once we find the color we're looking for, we'll return the ColorInfo for that node (which is the value at the root without either of it's sub-trees).

Be sure to write a specification for the function. Write your specification with tags such as @returns, rather than rewriting the above formal definition (which is more information than the client wants or needs).

(d) Write test cases for lookup in the file color_tree_test.ts. There is one example test given., Follow the rules taught in lecture for choosing appropriate test cases.

Include brief comments in your test suite justifying the test cases you chose.

Confirm that all your tests pass by running npm run test.

(e) We have everything we need to start using the tree to lookup color names. In order to use it in the rest of the code, however, we need to implement the ColorList interface that we created earlier.

Create a class in color_tree.ts called ColorTree that implements the ColorList interface. It should take a List<ColorInfo> and store it in a field as well as store the same data in tree form as created by calling makeBst.

Be sure to properly document the class, which should now have both an AF and RI.

There is no efficiency benefit to using the tree to implement findMatchingNames, so instead, you should call the existing function findMatchingNamesIn from color_list.ts (which you must now edit to export), passing it the list of colors instead of the tree.

Implement the function getColorCss by calling lookup on the tree. (See the getColorCssIn function in color_list.ts for an example of what to return if the color is found and what to throw if it is not found). Also, note that you may need to explicitly include the return type on the function you write, as shown here, in order to please the type checker:

```
getColorCss = (name: string): readonly [string, string] => {...}
```

Create a factory function, called makeColorTree, that returns an instance of ColorTree, passing it the colors from the COLORS list. As before, you should use the singleton pattern to avoid creating multiple instances. Be sure to write a specification for the function.

(f) We also need tests for the new code. However, since we are using the same interface as before, we can simply copy the existing tests for findMatchingNames and getColorCss from color_list_test.ts into color_tree_test.ts. The only change that should be needed is to call the new factory function, makeColorTree, instead of the old one.

Confirm that all the tests pass by running npm run test.

(g) Change the code in index.tsx to use makeColorTree instead of makeSimpleColorList.

Make sure that the app still works by running npm run start and trying it out.

More Lists

Earlier, we defined the list of numbers that appear in a tree. We also talked about whether a number appears in a list, but we never defined that precisely. We can do so as follows:

More **BSTs**

Earlier, we gave an informal description of what must be true in order to create node(x, S, T): every value in S must be less than or equal to x and every value in T must be greater than x. Formally, the BST invariant requires that the following must hold for any integer y:

- if contains(y, values(S)) = true, then $y \leq x$, and
- if contains(y, values(T)) = true, then y > x.

An equivalent way to say this is that, if y > x, then y cannot appear in S, and if $y \le x$, then y cannot appear in T. Formally, then, the BST invariant equivalently says that the following must hold for any integer y:

- if y > x, then contains(y, values(S)) = false, and
- if $y \le x$, then contains(y, values(T)) = false.

(Students who have taken 311 should recognize these statements as the contrapositives of those above.)

Another application of contains is that it will allow us to formally state and prove the correctness of lookup, which we will do in the next problem.

4. Many More Fish in the Tree (20 points)

The following parts consist entirely of written work. It should be submitted with "HW5 Written".

In this problem, we will prove the correctness of the BST lookup function we defined earlier. To get there, we first need to prove another fact.

(a) Let a be any integer and S be any list. Prove by induction on L that

contains(a, concat(L, S)) = contains(a, L) or contains(a, S)

Hint: It may be useful to experiment with calculating both left-to-right and right-to-left to get the equality to work. When you write your final proof, however you should only move from one side of the equation to the other.

(b) Let a be any integer. Prove by induction that, for any U : BST, we must have

 $contains(a, values(U)) = (lookup(a, U) \neq undefined)$

In other words, lookup returns undefined when contains(a, values(U)) = false, and it returns a node in the tree when contains(a, values(U)) = true.

Hints for the inductive step:

Start by proving, by calculation, and using part (a), that

contains(a, values(node(b, S, T))) = contains(a, values(S)) or (a = b) or contains(a, values(T))

• Then, continue the argument by cases over whether a = b, a < b, or a > b. In each case, you should be able to prove that the right-hand side is lookup(a, node(b, S, T)). In the case a < b, for example, you should be able to show that the right-hand side simplifies to contains(a, values(S)), using the BST invariant, and then show that the simplified expression equals lookup(a, node(b, S, T)).

5. Chomping at the Split (20 points)

The following parts consist of a mix of written and coding work: part (b) is written and should be submitted with "HW5 Written", while parts (a,c,d,e,f) are coding and should be submitted with "HW5 Coding".

In this problem we will change the format in which users write the text that they want highlighted. Instead of asking them to split the text into lines, with each line having text in a single color, we will allow them to write all the text in one line and indicate where they want highlighting using the syntax "[color|text]", where "color" is the name of the color and "text" is everything they want shown in that color. All text not written inside "[..]" will be written without a highlight (i.e., black text on a white background). For example:

[MediumPurple|hello] [CornflowerBlue|,] [LightSeaGreen|world] [DarkOrange|!]



In order to do that, we will first need a helper function, split-at, that takes a list L and a value c as arguments and "returns a pair of lists (P, S), with L = concat(P, S), where P contains all the values before the first c and S is either empty or contains the remainder of the values starting with the first c". For example, the calling split-at with inputs [2, 8, 3, 4, 3] and 3 would return ([2, 8], [3, 4, 3]).

The following is a formal definition of split-at using recursion:

 $\begin{array}{lll} \mbox{func split-at(nil, c)} & := & (nil, nil) & \mbox{for any } c: \mathbb{Z} \\ \mbox{split-at}(\mbox{cons}(a, R), c) & := & (nil, \mbox{cons}(a, R)) & \mbox{if } a = c & \mbox{for any } a, c: \mathbb{Z} \mbox{ and } R: \mbox{List} \\ \mbox{split-at}(\mbox{cons}(a, R), c) & := & (\mbox{cons}(a, P), S) & \mbox{if } a \neq c & \mbox{for any } a, c: \mathbb{Z} \mbox{ and } R: \mbox{List} \\ \mbox{where } (P, S) = \mbox{split-at}(R, c) \end{array}$

A translation of this function in to Typescript is given in the list.ts file with the following signature:

export const split_at = <A,>(L: List<A>, c: A): readonly [List<A>, List<A>] => {..};

(a) Write test cases for split_at in list_test.ts. Follow the rules taught in lecture for choosing appropriate test cases.

Include brief comments in your test suite justifying the test cases you chose.

Confirm that all your tests pass by running npm run test.

To make parsing easier, we have provided you with a helper function getNextHighlight that finds the next highlighted portion of the text, if one exists. It returns either a triple (b, h, a), where b is the text before the next highlight (as a *string*), h is the highlight that was found (a record containing the color name and text), and a is the text after the next highlight (as a *List*), or undefined if there is no next highlighted portion.

Note that this function parses the *first* highlight found in the text, but not anything after that. In other words, when it returns a triple (b, h, a), while the string b will not contain any text of the form "[color|text]", the list a very well might, so more work is necessary to fully parse the text.

Our goal is to use the helper function just described to write a function that parses a list of characters into a list of highlight records that describes all the text. This includes not only the portions of the form "[color|text]", where the user indicated a specific color, but also all the regular text, which should be turned into records with background color "white". For example, the text "my [red|favorite] book" would become the list

cons({color: "white", text: "my "}, cons({color: "red", text: "favorite"}, cons({color: "white", text: " book"},nil))) Be careful to ensure that the list you return does not include any extra, empty highlight records (i.e., records where the text field is the empty string) that are not actually described in the input.

For example, if the input text is "[red|one][blue|two]", your result should contain two highlight records. There shouldn't be an extra {color: "white", text: ""} record before, after, or between the two highlights. On the other hand, if the user actually writes "[red|]", then they are explicitly describing a red highlight of the empty string, so you should include that in your answer.

(b) We were given an English definition of the problem, so our first step is to formalize it.

Write a formal definition of parse using calls to getNextHighlight, which we will abbreviate "nh". You are also free to use the "compact" function that turns a list back into a string.

(Reminder: doing this on paper — trying to get all the cases right without testing it out on a computer — is good practice for interviews, where you will not have computer help!)

(c) Translate your definition of parse into a TypeScript function in parser.ts with the following signature:

export const findHighlights = (chars: List<number>): List<Highlight> => {..};

The type Highlight is the record type used above, with fields called "color" and "text", both strings.

(d) Write test cases for findHighlights in the file parse_test.ts.

Include brief comments in your test suite justifying the test cases you chose.

Confirm that all your tests pass by running npm run test.

(e) The starter code includes a simple wrapper function of findHighlights called parseHighlightText. It takes input as a string and converts it into a list before calling findHighlights.

Uncomment the code for parseHighlightText in parser.ts and the tests for it in parser_test.ts. Confirm these tests pass by running npm run test.

(f) Change the code in index.tsx to use parseHighlightText instead of parseHighlightLines. Make sure that the app now supports the new syntax by running npm run start and trying it out. Congratulations! You've made the app both faster and more useful. Great work!

6. Extra Credit: Live Tree or Die (0 points)

The following parts consist of entirely written work. It should be submitted with "HW5 Written". On the previous page, we claimed that we have made the app faster by replacing the list search with a BST search, but we haven't actually proven that. We will do that in this extra credit problem.

In order to do so, we need the following definition:

With that in hand, we can now describe the running time of lookup and show that it is faster than linear.

- (a) Let L be any list. Show that $len(P) \le len(S)$, where (P, S) = split(m, L) and m = |len(L)/2|.
- (b) Suppose that $len(L) \le 2^{k+1} 1$ for some integer k. Show that we then have $len(P) \le 2^k 1$, where P is defined as in part (a).(You may need to use part (a) to show this.)
- (c) Let L be any non-empty list, and let a any integer. Show that we have $len(R) \le len(P)$, where (P, S) = split(m, cons(a, L)), $m = \lfloor len(cons(a, L))/2 \rfloor$, and S = cons(b, R).

It may be easiest to argue by cases based on whether n = len(cons(a, L)) is even or odd.

- (d) Let $a, b, c : \mathbb{Z}$ with $a \le c$ and $b \le c$. Prove by cases that $\max(a, b) \le c$.
- (e) Prove by induction that, for any list L, the inequality height(makeBst(L)) $\leq k$ holds for any integer k satisfying len(L) $\leq 2^{k+1} 1$.

You may need to use parts (b-d).

- (f) Explain how to prove that, for any non-empty tree T and integer y, the number of recursive calls made by lookup(y,T) before it returns (not just the first recursive call but any recursive calls that makes and so on) is no more than height(T). (You do not need to write out the proof in full detail.)
- (g) Let L be a list and n := len(L). The integer $k := \lceil \log_2(n+1) \rceil 1$, where $\lceil \cdot \rceil$ means rounding up to the closest larger integer, satisfies $n \le 2^{k+1} 1$.

Explain why we know that lookup(y,T), for any integer y, runs in O(log n) time. (This is exponentially faster than linear search, which runs in $\Theta(n)$ time.)