Technische Universität München
Institut für Informatik
Prof. Tobias Nipkow, Ph.D.
Fabian Huch

SS 2021
4. 6. 2021

Functional Data Structures

Exercise Sheet 8

Exercise 8.1 Joining 2-3-Trees

Implement and prove correct a function to combine two 2-3-trees of equal height, such that the inorder traversal of the resulting tree is the concatenation of the inorder traversal of the arguments, and the height of the result is either the height of the arguments, or has increased by one. Use *'a upI* to return the result, similar to *Tree23_Set.ins*:

fun *join* :: "'*a tree23* ⇒ '*a tree23* ⇒ '*a upI*"

lemma *join_inorder*:
  fixes *t1 t2* :: "'*a tree23*"
  assumes "*height t1* = *height t2*"
  assumes "*complete t1*" "*complete t2*"
  shows "*inorder* (*treeI* (*join t1 t2*)) = (*inorder t1* @ *inorder t2*)"

lemma *join_complete*:
  fixes *t1 t2* :: "'*a tree23*"
  assumes "*height t1* = *height t2*"
  assumes "*complete t1*" "*complete t2*"
  shows "*complete* (*treeI* (*join t1 t2*)) ∧ *hI* (*join t1 t2*) = *height t2*"

Hints:

- Try to use automatic case splitting (*auto split*: ...) instead of explicit case splitting via Isar (There will be dozens of cases).
- To find bugs in your join function, or isolate the case where your automatic proof does not (yet) work, use Isar to perform the induction proof case by case.

Exercise 8.2 Bounding the Size of 2-3-Trees

Show that for 2-3-trees, we have:

$$log_3\ (\ s(t)\ +\ 1\ ) \leq h(t) \leq log_2\ (s(t)\ +\ 1)$$

Hint: It helps to first raise the two sides of the inequation to the 2nd/3rd power. Use sledgehammer and find-theorems to search for the appropriate lemmas.

lemma *height_bound_upper*: "$complete\ t \Longrightarrow height\ t \leq log\ 2\ (size\ t + 1)$"

lemma *height_bound_lower*:
assumes "$complete\ t$"
shows "$log\ 3\ (size\ t + 1) \leq height\ t$"

Homework 8.1 Fibonacci tree is a minimal AVL tree

Submission until Thursday, June 10, 23:59pm.

Take a look at the following definition of fibonacci trees (which store only the height):

fun *fib_tree* :: "$nat \Rightarrow nat\ tree$" **where**
"$fib\_tree\ 0 = \langle\rangle$" |
"$fib\_tree\ (Suc\ 0) = \langle\langle\rangle,\ 1,\ \langle\rangle\rangle$" |
"$fib\_tree\ (Suc(Suc\ n)) = \langle fib\_tree\ (Suc\ n),\ Suc\ (Suc\ n),\ fib\_tree\ n\rangle$"

Interestingly, the trees constructed by the above function represent a lower bound on the compactness of AVL trees in following sense: a Fibonacci tree of a given height has at least the same number of inner nodes as that of an AVL tree of the same height. Prove this law (you can find a proof sketch in the book).

lemma *fib_tree_minimal*: "$avl\ t \Longrightarrow size1\ (fib\_tree\ (ht\ t)) \leq size1\ t$"

Homework 8.2 Deletion from a disjoint interval tree

Submission until Thursday, June 10, 23:59pm.

An interval tree is a tree whose nodes each contain an interval of elements from an ordered type:

datatype $'a\ itree = iLeaf \mid iNode$ "$'a\ itree$" "$'a \times 'a$" "$'a\ itree$"

The following are two useful functions for interval trees: one returning the set of intervals in the tree, and another returning the set of elements in the tree.

fun *set_itree2*:: "$'a::ord\ itree \Rightarrow 'a\ set$" **where**
"$set\_itree2\ iLeaf = \{\}$"
| "$set\_itree2\ (iNode\ l\ (low,\ high)\ r) = \{low\ ..\ high\} \cup ((set\_itree2\ l) \cup (set\_itree2\ r))$"

fun *set_itree3*:: "$'a\ itree \Rightarrow ('a \times 'a)\ set$" **where**
"$set\_itree3\ iLeaf = \{\}$"
| "$set\_itree3\ (iNode\ l\ (low,\ high)\ r) = \{(low,\ high)\} \cup ((set\_itree3\ l) \cup (set\_itree3\ r))$"

An ordered disjoint interval tree is an interval tree such that:

- The lower end of an interval in a node is strictly greater than the higher end of every interval in the left subtree.
- The higher end of an interval in a node is strictly smaller than the lower end in every interval in the right subtree.
- The interval in every node has a lower end that is smaller than or equal than its upper end.

Recursively define an invariant for an interval tree that formalises the above conditions.

fun *ibst* :: *“'a::linorder itree ⇒ bool”*

Define and verify a delete function for interval trees. That function should: i) only take an element (i.e. not an interval) and delete it from the tree, ii) exploit the fact that the tree is ordered, and iv) be implemented using an appropriate join function for interval trees.

fun *delete* :: *“int ⇒ int itree ⇒ int itree”*

Hint: this function has to deal with three cases.

- if the element is equal to the two ends of an interval, in which case the interval should be completely removed from the tree,
- if the element is equal to one of the ends of an interval, in which case the interval has to be appropriately shrinked, and
- if the element lies with an interval, in which case the interval as to be split into two. One way to deal with that situation is to let the left subinterval inherit the position of the original interval, and position the right subinterval to be the left most leaf in the right subtree.

Prove that the function removes the correct element from the tree.

theorem *delete_set_minus*: *“ibst t* $\Longrightarrow$ *set_itree2 (delete x t) = (set_itree2 t)* $-$ $\{x\}$*”*

Prove that the resulting interval tree conforms to the invariant. Hint: you might want to define a function that returns an list of intervals in an interval tree and a predicate characterising the sortedness of that list. Proving that *delete* preserves the invariant should reduce to arguing about the sortedness of this list.

theorem *delete_ibst*: *“ibst t* $\Longrightarrow$ *ibst (delete x t)”*

Homework 8.3 Size of 2-Trees (bonus)

Submission until Thursday, June 10, 23:59pm.

This is a bonus exercise worth 3 points. Sledgehammer-proofs (i.e., using *metis*, *smt*, *meson*, or *moura*) are not allowed.

Define a function to count the number of leaves of a 2-3-tree, and a function to determine whether a tree only consists of 2-nodes and leaves.

fun *num_leaves* :: "'a tree23 ⇒ nat"
fun *is_2_tree* :: "'a tree23 ⇒ bool"

Show that a 2-3-tree has only 2 nodes, if and only if its number of leafs is 2 to the power of its height!

theorem *complete_2_tree_height*: "complete t ⟹ is_2_tree t ⟷ num_leaves t = 2^height t"