Functional Programming
Practicals 2: Red-Black Tree

Shin-Cheng Mu

FLOLAC 2022

In this practical we aim to prove some essential properties about red-black tree insertion in order to establish the correctness of the insertion algorithm. Some notes:

- In most proofs there could be many repetitive cases. It is sufficient to show only some representative cases.
- Proof about properties of the function *balance* are mostly routine, tedious, non-inductive proofs. However, these properties are needed in other proofs.

The code are adapted from Okasaki [?]. Those who interested in figuring out how to perform deletion in red-black trees may check out Germane and Might [?].

1. Complete the definitions in the file `RedBlackOkasaki.hs`.
2. On (black) heights.
   (a) Prove that forall t, u and z, $\mathit{bheight}\ (\mathit{balance}\ t\ z\ u) = 1 + (\mathit{bheight}\ t \uparrow \mathit{bheight}\ u)$.

Solution: We need only a non-inductive proof that checks all the cases. We demonstrate only one of them.

Case $(t, z, u) := (\mathrm{R}\ (\mathrm{R}\ t\ x\ u)\ y\ v, z, w)$:

$$
\begin{array}{ll}
& \mathit{bheight}\ (\mathit{balance}\ (\mathrm{R}\ (\mathrm{R}\ t\ x\ u)\ y\ v)\ z\ w) \\
= & \quad \{ \text{ def. of } \mathit{balance} \ \} \\
& \mathit{bheight}\ (\mathrm{R}\ (\mathrm{B}\ t\ x\ u)\ y\ (\mathrm{B}\ v\ z\ w)) \\
= & \quad \{ \text{ def. of } \mathit{bheight} \ \} \\
& \mathit{bheight}\ (\mathrm{B}\ t\ x\ u) \uparrow \mathit{bheight}\ (\mathrm{B}\ v\ z\ w) \\
= & \quad \{ \text{ def. of } \mathit{bheight} \ \} \\
& (1 + (\mathit{bheight}\ t \uparrow \mathit{bheight}\ u)) \uparrow \\
& (1 + (\mathit{bheight}\ v \uparrow \mathit{bheight}\ w)) \\
= & \quad \{ \text{ since } (k + x) \uparrow (k + y) = k + (x \uparrow y), (\uparrow) \text{ associative } \} \\
& 1 + (((\mathit{bheight}\ t \uparrow \mathit{bheight}\ u) \uparrow \mathit{bheight}\ v) \uparrow \mathit{bheight}\ w) \\
= & \quad \{ \text{ definition of } \mathit{bheight} \ \} \\
& 1 + (\mathit{bheight}\ ((\mathrm{R}\ (\mathrm{R}\ t\ x\ u)\ y\ v)) \uparrow \mathit{bheight}\ w)
\end{array}
$$

(b) Prove that for all k and t, $bheight\ (ins\ k\ t) = bheight\ t$.
Note: as a corollary, we have $bheight\ (insert\ k\ t)$ equals either $bheight\ t$ or $1 + bheight\ t$, depending on the root color of $ins\ k\ t$.

Solution: Induction on n. We show only some representative cases.

Case $t := \mathrm{E}$.

$$
\begin{aligned}
& bheight\ (ins\ k\ \mathrm{E}) \\
=\ & bheight\ (\mathrm{R\ E}\ k\ \mathrm{E}) \\
=\ & 0 \\
=\ & bheight\ \mathrm{E}\ \ .
\end{aligned}
$$

Case $t := \mathrm{R}\ t\ x\ u$, $k < x$:

$$
\begin{aligned}
& bheight\ (ins\ k\ (\mathrm{R}\ t\ x\ u)) \\
=\ & \quad \{ \text{ def. of } ins,\ k < x \ \} \\
& bheight\ (\mathrm{R}\ (ins\ k\ t)\ x\ u) \\
=\ & \quad \{ \text{ def. of } bheight \ \} \\
& bheight\ (ins\ k\ t) \uparrow bheight\ u \\
=\ & \quad \{ \text{ induction } \} \\
& bheight\ t \uparrow bheight\ u \\
=\ & \quad \{ \text{ def. of } bheight \ \} \\
& bheight\ (\mathrm{R}\ t\ x\ u)\ \ .
\end{aligned}
$$

Case $t := \mathrm{B}\ t\ x\ u$, $k < x$:

$$
\begin{aligned}
& bheight\ (ins\ k\ (\mathrm{B}\ t\ x\ u)) \\
=\ & \quad \{ \text{ def. of } ins,\ k < x \ \} \\
& bheight\ (balance\ (ins\ k\ t)\ x\ u) \\
=\ & \quad \{ \text{ exercise 2(a) } \} \\
& 1 + (bheight\ (ins\ k\ t) \uparrow bheight\ u) \\
=\ & \quad \{ \text{ induction } \} \\
& 1 + (bheight\ t \uparrow bheight\ u) \\
=\ & \quad \{ \text{ def. of } bheight \ \} \\
& bheight\ (\mathrm{B}\ t\ x\ u)\ \ .
\end{aligned}
$$

3. On balancing.

(a) The function *isBalanced*, when taken literally as an algorithm, has time complexity $O(n^2)$, where n is the size of the input tree. Define

$$
\begin{aligned}
& isBalHeight :: \mathrm{RBTree}\ a \rightarrow (\mathrm{Bool}, \mathrm{Nat}) \\
& isBalHeight\ t = (isBalanced\ t, bheight\ t)\ \ .
\end{aligned}
$$

Derive an implementation of *isBalHeight* that runs in time linear to the size of the input tree.

Solution: Do a case analysis on t. When $t := \mathrm{E}$ we clearly have $\mathit{isBalHeight}\ \mathrm{E} = (\mathrm{True}, 0)$.

For $t := \mathrm{B}\ t\ x\ u$, we calculate:

$$
\begin{array}{ll}
 & \mathit{isBalHeight}\ (\mathrm{B}\ t\ x\ u) \\
= & \quad \{\ \text{def. of } \mathit{isBalHeight}\ \} \\
 & (\mathit{isBalanced}\ (\mathrm{B}\ t\ x\ u), \mathit{bheight}\ (\mathrm{B}\ t\ x\ u)) \\
= & \quad \{\ \text{def. of } \mathit{isBalanced} \text{ and } \mathit{bheight}\ \} \\
 & (\mathit{bheight}\ t == \mathit{bheight}\ u \wedge \mathit{isBalanced}\ t \wedge \mathit{isBalanced}\ u, \\
 & \quad 1 + (\mathit{bheight}\ t \uparrow \mathit{bheight}\ u)) \\
= & \quad \{\ \text{grouping calls to } \mathit{isBalanced} \text{ and } \mathit{bheight} \text{ together}\ \} \\
 & \mathbf{let}\ (\mathit{bt}, \mathit{ht}) = (\mathit{isBalanced}\ t, \mathit{bheight}\ t) \\
 & \qquad (\mathit{bu}, \mathit{hu}) = (\mathit{isBalanced}\ u, \mathit{bheight}\ u) \\
 & \mathbf{in}\ \ (\mathit{ht} == \mathit{hu} \wedge \mathit{bt} \wedge \mathit{bu}, 1 + (\mathit{ht} \uparrow \mathit{hu})) \\
= & \quad \{\ \text{def. of } \mathit{isBalHeight}\ \} \\
 & \mathbf{let}\ (\mathit{bt}, \mathit{ht}) = \mathit{isBalHeight}\ t \\
 & \qquad (\mathit{bu}, \mathit{hu}) = \mathit{isBalHeight}\ u \\
 & \mathbf{in}\ \ (\mathit{ht} == \mathit{hu} \wedge \mathit{bt} \wedge \mathit{bu}, 1 + (\mathit{ht} \uparrow \mathit{hu}))\ \ .
\end{array}
$$

The case for $t := \mathrm{R}\ t\ x\ u$ is similar. In summary we have:

$$
\begin{array}{l}
\mathit{isBalHeight} :: \mathrm{RBTree}\ a \rightarrow (\mathrm{Bool}, \mathrm{Nat}) \\
\mathit{isBalHeight}\ \mathrm{E} = (\mathrm{True}, 0) \\
\mathit{isBalHeight}\ (\mathrm{R}\ t\ x\ u) = \\
\quad \mathbf{let}\ (\mathit{bt}, \mathit{ht}) = \mathit{isBalHeight}\ t \\
\qquad\quad (\mathit{bu}, \mathit{hu}) = \mathit{isBalHeight}\ u \\
\quad \mathbf{in}\ \ (\mathit{ht} == \mathit{hu} \wedge \mathit{bt} \wedge \mathit{bu}, (\mathit{ht} \uparrow \mathit{hu})) \\
\mathit{isBalHeight}\ (\mathrm{B}\ t\ x\ u) = \\
\quad \mathbf{let}\ (\mathit{bt}, \mathit{ht}) = \mathit{isBalHeight}\ t \\
\qquad\quad (\mathit{bu}, \mathit{hu}) = \mathit{isBalHeight}\ u \\
\quad \mathbf{in}\ \ (\mathit{ht} == \mathit{hu} \wedge \mathit{bt} \wedge \mathit{bu}, 1 + (\mathit{ht} \uparrow \mathit{hu}))\ \ .
\end{array}
$$

(b) Prove that for all t and u,

$$
\begin{array}{l}
\mathit{isBalanced}\ t \wedge \mathit{isBalanced}\ u\ \wedge \\
\quad \mathit{bheight}\ t = \mathit{bheight}\ u \Rightarrow \mathit{isBalanced}\ (\mathit{balance}\ t\ x\ u)\ \ .
\end{array}
$$

Solution: A tedious but routine check. We show only one of the cases.

Case: $(t, x, u) := (\mathsf{R}\ (\mathsf{R}\ t\ x\ u)\ y\ v, z, w)$.

$$
\begin{aligned}
& \mathit{isBalanced}\ (\mathit{balance}\ (\mathsf{R}\ (\mathsf{R}\ t\ x\ u)\ y\ v)\ z\ w) \\
= & \quad \{ \text{ def. of } \mathit{balance} \ \} \\
& \mathit{isBalanced}\ (\mathsf{R}\ (\mathsf{B}\ t\ x\ u)\ y\ (\mathsf{B}\ v\ z\ w)) \\
= & \quad \{ \text{ def. of } \mathit{isBalanced} \ \} \\
& \mathit{bheight}\ (\mathsf{B}\ t\ x\ u) = \mathit{bheight}\ (\mathsf{B}\ v\ z\ w) \wedge \\
& \mathit{isBalanced}\ (\mathsf{B}\ t\ x\ u) \wedge \mathit{isBalanced}\ (\mathsf{B}\ v\ z\ w) \\
= & \quad \{ \text{ def. of } \mathit{bheight} \ \} \\
& 1 + (\mathit{bheight}\ t \uparrow \mathit{bheight}\ u) = 1 + (\mathit{bheight}\ v \uparrow \mathit{beight}\ w) \wedge \\
& \mathit{bheight}\ t = \mathit{bheight}\ u \wedge \mathit{isBalanced}\ t \wedge \mathit{isBalanced}\ u \wedge \\
& \mathit{bheight}\ v = \mathit{bheight}\ w \wedge \mathit{isBalanced}\ v \wedge \mathit{isBalanced}\ w \\
\Leftarrow & \quad \{ \text{ def. of } \mathit{isBalanced} \text{ and } \mathit{bheight}\text{, arithmetics } \} \\
& \mathit{bheight}\ (\mathsf{R}\ t\ x\ u) = \mathit{bheight}\ v \wedge \\
& \mathit{isBalanced}\ (\mathsf{R}\ t\ x\ u) \wedge \mathit{isBalanced}\ v \wedge \mathit{isBalanced}\ w \wedge \\
& \mathit{bheight}\ t \uparrow \mathit{bheight}\ u \uparrow \mathit{bheight}\ v = \mathit{bheight}\ w \\
= & \quad \{ \text{ def. of } \mathit{isBalanced} \text{ and } \mathit{bheight} \ \} \\
& \mathit{isBalanced}\ (\mathsf{R}\ (\mathsf{R}\ t\ x\ u)\ y\ v) \wedge \mathit{isBalanced}\ w \wedge \\
& \quad \mathit{bheight}\ \mathsf{R}\ (\mathsf{R}\ t\ x\ u)\ y\ v = \mathit{bheight}\ w \ \ .
\end{aligned}
$$

(c) Prove that for all k and t, $\mathit{isBalanced}\ t \Rightarrow \mathit{isBalanced}\ (\mathit{ins}\ k\ t)$.
Note: since $\mathit{isBalanced}\ t \Rightarrow \mathit{isBalanced}\ (\mathit{blacken}\ t)$, as a corollary we have $\mathit{isBalanced}\ t \Rightarrow \mathit{isBalanced}\ (\mathit{insert}\ k\ t)$.

Solution: Induction on t. The base case $t := \mathsf{E}$ is omitted. We demonstrate one of the cases.

Case $t := \mathsf{B}\ t\ x\ u$, $k < x$:

$$
\begin{aligned}
& \mathit{isBalanced}\ (\mathit{ins}\ k\ (\mathsf{B}\ t\ x\ u)) \\
= & \quad \{ \text{ def. of } \mathit{ins}\text{, } k < x \ \} \\
& \mathit{isBalanced}\ (\mathit{balance}\ (\mathit{ins}\ k\ t)\ x\ u) \\
\Leftarrow & \quad \{ \text{ exercise 3(b) } \} \\
& \mathit{isBalanced}\ (\mathit{ins}\ k\ t) \wedge \mathit{isBalanced}\ u \wedge \mathit{bheight}\ (\mathit{ins}\ k\ t) = \mathit{bheight}\ u \\
= & \quad \{ \text{ exercise 2(b) } \} \\
& \mathit{isBalanced}\ (\mathit{ins}\ k\ t) \wedge \mathit{isBalanced}\ u \wedge \mathit{bheight}\ t = \mathit{bheight}\ u \\
\Leftarrow & \quad \{ \text{ induction } \} \\
& \mathit{isBalanced}\ t \wedge \mathit{isBalanced}\ u \wedge \mathit{bheight}\ t = \mathit{bheight}\ u \\
= & \quad \{ \text{ def. of } \mathit{isBalanced} \ \} \\
& \mathit{isBalanced}\ (\mathsf{B}\ t\ x\ u) \ \ .
\end{aligned}
$$

4. On color invariants.

(a) Prove that for all t and u, $\mathit{islRB}\ t \wedge \mathit{isRB}\ u \Rightarrow \mathit{isRB}\ (\mathit{balance}\ t\ x\ u)$.

Solution: Note that with the constraint *isRB u*, we need to check only the 1st, 2nd, and last case of *balance*. The proof is rather straight forward. Take for example the 1st case:

Case: $(t, x, u) := (\mathrm{R}\ (\mathrm{R}\ t\ x\ u)\ y\ v, z, w)$:

$$\begin{array}{cl}
& \mathit{isRB}\ (\mathit{balance}\ (\mathrm{R}\ (\mathrm{R}\ t\ x\ u)\ y\ v)\ z\ w) \\
= & \quad \{ \text{def. of } \mathit{balance} \} \\
& \mathit{isRB}\ (\mathrm{R}\ (\mathrm{B}\ t\ x\ u)\ y\ (\mathrm{B}\ v\ z\ w)) \\
= & \quad \{ \text{def. of } \mathit{isRB} \} \\
& \mathit{color}\ (\mathrm{B}\ t\ x\ u) = \mathrm{Blk} \wedge \mathit{color}\ (\mathrm{B}\ v\ z\ w) = \mathrm{Blk} \wedge \\
& \mathit{isRB}\ (\mathrm{B}\ t\ x\ u) \wedge \mathit{isRB}\ (\mathrm{B}\ v\ z\ w) \\
= & \quad \{ \text{def. of } \mathit{color} \} \\
& \mathit{isRB}\ (\mathrm{B}\ t\ x\ u) \wedge \mathit{isRB}\ (\mathrm{B}\ v\ z\ w) \\
= & \quad \{ \text{def. of } \mathit{isRB} \} \\
& \mathit{color}\ t = \mathit{color}\ u = \mathit{color}\ v = \mathit{color}\ w = \mathrm{Blk} \wedge \\
& \mathit{isRB}\ t \wedge \mathit{isRB}\ u \wedge \mathit{isRB}\ v \wedge \mathit{isRB}\ w \\
= & \quad \{ \text{def. of } \mathit{isRB} \} \\
& \mathit{color}\ v = \mathit{color}\ w = \mathrm{Blk} \wedge \\
& \mathit{isRB}\ (\mathrm{R}\ t\ x\ u) \wedge \mathit{isRB}\ v \wedge \mathit{isRB}\ w \\
= & \quad \{ \text{def. of } \mathit{islRB} \} \\
& \mathit{islRB}\ (\mathrm{R}\ (\mathrm{R}\ t\ x\ u)\ y\ v) \wedge \mathit{isRB}\ w\ \ .
\end{array}$$

(b) Prove that for all t:

1. $\mathit{isRB}\ t \wedge \mathit{color}\ t = \mathrm{R} \Rightarrow \mathit{islRB}\ (\mathit{ins}\ k\ t)$,
2. $\mathit{isRB}\ t \wedge \mathit{color}\ t = \mathrm{B} \Rightarrow \mathit{isRB}\ (\mathit{ins}\ k\ t)$.

Hints: 1. The two properties shall be proved simultaneously in one inductive proof. 2. Since $\mathit{isRB}\ t \Rightarrow \mathit{islRB}\ t$, the two properties above imply that $\mathit{isRB}\ t \Rightarrow \mathit{islRB}\ (\mathit{ins}\ k\ t)$, which you may need in the proof.

Note: since $\mathit{islRB}\ t \Rightarrow \mathit{isRB}\ (\mathit{blacken}\ t)$, as a corollary we have $\mathit{isRB}\ t \Rightarrow \mathit{isRB}\ (\mathit{insert}\ k\ t)$.

Solution: Induction on t. We demonstrate two cases:

Case $t := \mathrm{B}\ t\ x\ u$, $k < x$.

$$\begin{array}{cl}
& \mathit{isRB}\ (\mathit{ins}\ k\ (\mathrm{B}\ t\ x\ u)) \\
= & \quad \{ \text{def. of } \mathit{ins}, k < x \} \\
& \mathit{isRB}\ (\mathit{balance}\ (\mathit{ins}\ k\ t)\ x\ u) \\
\Leftarrow & \quad \{ \text{exercise 4(a)} \}
\end{array}$$

$$
\begin{array}{ll}
& \mathit{isIRB}\,(\mathit{ins}\ k\ t) \wedge \mathit{isRB}\ u \\
\Leftarrow & \{\ \text{induction, noting that } \mathit{isRB}\ t \Rightarrow \mathit{isIRB}\,(\mathit{ins}\ k\ t)\ \} \\
& \mathit{isRB}\ t \wedge \mathit{isRB}\ u \\
= & \{\ \text{def. of } \mathit{isRB}\ \} \\
& \mathit{isRB}\,(\mathrm{B}\ t\ x\ u)\ \ .
\end{array}
$$

Case $t := \mathrm{R}\ t\ x\ u$, $k < x$.

$$
\begin{array}{ll}
& \mathit{isIRB}\,(\mathit{ins}\ k\ (\mathrm{R}\ t\ x\ u)) \\
= & \{\ \text{def. of } \mathit{ins}\ \} \\
& \mathit{isRB}\,(\mathrm{R}\ (\mathit{ins}\ k\ t)\ x\ u) \\
= & \{\ \text{def. of } \mathit{isIRB}\ \} \\
& (\mathit{color}\ (\mathit{ins}\ k\ t) = \mathrm{Blk} \vee \mathit{color}\ u = \mathrm{Blk}) \wedge \mathit{isRB}\,(\mathit{ins}\ k\ t) \wedge \mathit{isRB}\ u \\
= & \{\ \text{induction}\ \} \\
& (\mathit{color}\ (\mathit{ins}\ k\ t) = \mathrm{Blk} \vee \mathit{color}\ u = \mathrm{Blk}) \wedge \mathit{isRB}\ t \wedge \mathit{isRB}\ u \\
\Leftarrow & \{\ \text{logic: } ((\mathrm{P} \mid \mathrm{Q}) \wedge \mathrm{R}) \Leftarrow (\mathrm{Q} \wedge \mathrm{R})\ \} \\
& \mathit{color}\ u = \mathrm{Blk} \wedge \mathit{isRB}\ t \wedge \mathit{isRB}\ u \\
\Leftarrow & \{\ \text{logic: } \mathrm{P} \leqslant \mathrm{P} \wedge \mathrm{Q}\ \} \\
& \mathit{color}\ t = \mathit{color}\ u = \mathrm{Blk} \wedge \mathit{isRB}\ t \wedge \mathit{isRB}\ u \\
= & \{\ \text{def. of } \mathit{isRB}\ \} \\
& \mathit{isRB}\,(\mathrm{R}\ t\ x\ u)\ \ .
\end{array}
$$