

WIKIPEDIA
The Free Encyclopedia

Red–black tree

In computer science, a **red–black tree** is a specialised binary search tree data structure noted for fast storage and retrieval of ordered information, and a guarantee that operations will complete within a known time. Compared to other self-balancing binary search trees, the nodes in a red–black tree hold an extra bit called "color" representing "red" and "black" which is used when re-organising the tree to ensure that it is always approximately balanced.^[3]

When the tree is modified, the new tree is rearranged and "repainted" to restore the coloring properties that constrain how unbalanced the tree can become in the worst case. The properties are designed such that this rearranging and recoloring can be performed efficiently.

The (re-)balancing is not perfect, but guarantees searching in Big O time of $O(\log n)$, where n is the number of entries (or keys) in the tree. The insert and delete operations, along with the tree rearrangement and recoloring, are also performed in $O(\log n)$ time.^[4]

Tracking the color of each node requires only one bit of information per node because there are only two colors. The tree does not contain any other data specific to it being a red–black tree, so its memory footprint is almost identical to that of a classic (uncolored) binary search tree. In some cases, the added bit of information can be stored at no added memory cost.

Red–black tree		
Type	Tree	
Invented	1978	
Invented by Leonidas J. Guibas and Robert Sedgewick		
Complexities in big O notation		
Space complexity		
Space	$O(n)$	
Time complexity		
Function	Amortized	Worst Case
Search	$O(\log n)$ ^[1]	$O(\log n)$ ^[1]
Insert	$O(1)$ ^[2]	$O(\log n)$ ^[1]
Delete	$O(1)$ ^[2]	$O(\log n)$ ^[1]

History

In 1972, Rudolf Bayer^[5] invented a data structure that was a special order-4 case of a B-tree. These trees maintained all paths from root to leaf with the same number of nodes, creating perfectly balanced trees. However, they were not binary search trees. Bayer called them a "symmetric binary B-tree" in his paper and later they became popular as 2–3–4 trees or even 2–4 trees.^[6]

In a 1978 paper, "A Dichromatic Framework for Balanced Trees",^[7] Leonidas J. Guibas and Robert Sedgewick derived the red–black tree from the symmetric binary B-tree.^[8] The color "red" was chosen because it was the best-looking color produced by the color laser printer available to the authors while working at Xerox PARC.^[9] Another response from Guibas states that it was because of the red and black pens available to them to draw the trees.^[10]

In 1993, Arne Andersson introduced the idea of a right leaning tree to simplify insert and delete operations.^[11]

In 1999, Chris Okasaki showed how to make the insert operation purely functional. Its balance function needed to take care of only 4 unbalanced cases and one default balanced case.^[12]

The original algorithm used 8 unbalanced cases, but Cormen et al. (2001) reduced that to 6 unbalanced cases.^[3] Sedgewick showed that the insert operation can be implemented in just 46 lines of Java code.^{[13][14]} In 2008, Sedgewick proposed the left-leaning red–black tree, leveraging Andersson's idea that simplified the insert and delete operations. Sedgewick originally allowed nodes whose two children are red, making his trees more like 2–3–4 trees, but later this restriction was added, making new trees more like 2–3 trees. Sedgewick implemented the insert algorithm in just 33 lines, significantly shortening his original 46 lines of code.^{[15][16]}

Terminology

A red–black tree is a special type of binary search tree, used in computer science to organize pieces of comparable data, such as text fragments or numbers (as e. g. the numbers in figures 1 and 2). The nodes carrying keys and/or data are frequently called "internal nodes", but to make this very specific they are also called non-NIL nodes in this article.

The leaf nodes of red–black trees (NIL in figure 1) do not contain keys or data. These "leaves" need not be explicit individuals in computer memory: a NULL pointer can —as in all binary tree data structures— encode the fact that there is no child node at this position in the (parent) node. Nevertheless, by their position in the tree, these objects are in relation to other nodes that is relevant to the RB-structure, it may have parent, sibling (i. .e., the other child of the parent), uncle, even nephew node; and may be child—but never parent of another node. It is not really necessary to attribute a "color" to these end-of-path objects, because the condition "is NIL or BLACK" is implied by the condition "is NIL" (see also this remark).

Figure 2 shows the conceptually same red–black tree without these NIL leaves. To arrive at the same notion of a path, one must notice that e. g., 3 paths run through the node **1**, namely a path through **1**_{left} plus 2 added paths through **1**_{right}, namely the paths through **6**_{left} and **6**_{right}. This way, these ends of the paths are also docking points for new nodes to be inserted, fully equivalent to the NIL leaves of figure 1.

Instead, to save a marginal amount of execution time, these (possibly many) NIL leaves may be implemented as pointers to one unique (and black) sentinel node (instead of pointers of value NULL).

As a conclusion, the fact that a child does not exist (is not a true node, does not contain data) can in all occurrences be specified by the very same NULL pointer or as the very same pointer to a sentinel node. Throughout this article, either choice is called NIL node and has the constant value NIL.

The **black depth** of a node is defined as the number of black nodes from the root to that node (i. .e. the number of black ancestors). The **black height** of a red–black tree is the number of black nodes in any path from the root to the leaves, which, by requirement 4, is constant (alternatively, it

Example of a red–black tree

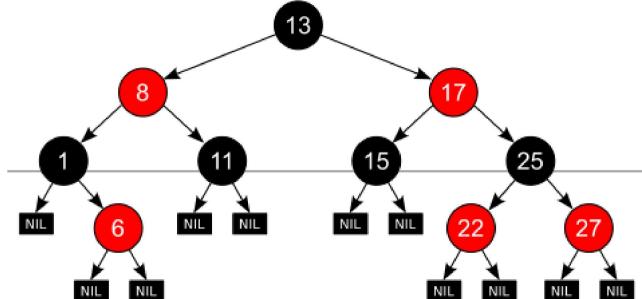


Figure 1: ... with explicit NIL leaves

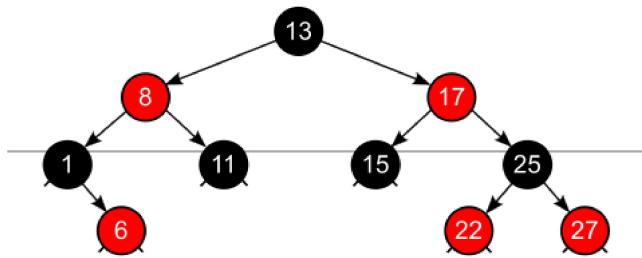


Figure 2: ... with implicit left and right docking points

could be defined as the black depth of any leaf node).^{[17]:154–165} The black height of a node is the black height of the subtree rooted by it. In this article, the black height of a NIL node shall be set to 0, because its subtree is empty as suggested by figure 2, and its tree height is also 0.

Properties

In addition to the requirements imposed on a binary search tree the following must be satisfied by a red–black tree:^[18]

1. Every node is either red or black.
2. All NIL nodes (figure 1) are considered black.
3. A red node does not have a red child.
4. Every path from a given node to any of its descendant NIL nodes goes through the same number of black nodes.
5. (Conclusion) If a node **N** has exactly one child, it must be a red child, because if it were black, its NIL descendants would sit at a different black depth than **N**'s NIL child, violating requirement 4.

Some authors, e. g. Cormen & al.,^[18] claim "the root is black" as fifth requirement; but not Mehlhorn & Sanders^[17] or Sedgewick & Wayne.^{[16]:432–447} Since the root can always be changed from red to black, this rule has little effect on analysis. This article also omits it, because it slightly disturbs the recursive algorithms and proofs.

As an example, every perfect binary tree that consists only of black nodes is a red–black tree.

The read-only operations, such as search or tree traversal, do not affect any of the requirements. In contrast, the modifying operations insert and delete easily maintain requirements 1 and 2, but with respect to the other requirements some extra effort must be made, to avoid introducing a violation of requirement 3, called a **red-violation**, or of requirement 4, called a **black-violation**.

The requirements enforce a critical property of red–black trees: *the path from the root to the farthest leaf is no more than twice as long as the path from the root to the nearest leaf*. The result is that the tree is height-balanced. Since operations such as inserting, deleting, and finding values require worst-case time proportional to the height ***h*** of the tree, this upper bound on the height allows red–black trees to be efficient in the worst case, namely logarithmic in the number ***n*** of entries, i. e. $h \in O(\log n)$, which is not the case for ordinary binary search trees. For a mathematical proof see section Proof of bounds.

Red–black trees, like all binary search trees, allow quite efficient sequential access (e. g. in-order traversal, that is: in the order Left–Root–Right) of their elements. But they support also asymptotically optimal direct access via a traversal from root to leaf, resulting in $O(\log n)$ search time.

Analogy to B-trees of order 4

A red–black tree is similar in structure to a B-tree of order^[19] 4, where each node can contain between 1 and 3 values and (accordingly) between 2 and 4 child pointers. In such a B-tree, each node will contain only one value matching the value in a black node of the red–black tree, with an optional value before and/or after it in the same node, both matching an equivalent red node of the red–black tree.

One way to see this equivalence is to "move up" the red nodes in a graphical representation of the red–black tree, so that they align horizontally with their parent black node, by creating together a horizontal cluster. In the B-tree, or in the modified graphical representation of the red–black tree, all leaf nodes are at the same depth.

The red–black tree is then structurally equivalent to a B-tree of order 4, with a minimum fill factor of 33% of values per cluster with a maximum capacity of 3 values.

This B-tree type is still more general than a red–black tree though, as it allows ambiguity in a red–black tree conversion—multiple red–black trees can be produced from an equivalent B-tree of order 4 (see figure 3). If a B-tree cluster contains only 1 value, it is the minimum, black, and has two child pointers. If a cluster contains 3 values, then the central value will be black and each value stored on its sides will be red. If the cluster contains two values, however, either one can become the black node in the red–black tree (and the other one will be red).

So the order-4 B-tree does not maintain which of the values contained in each cluster is the root black tree for the whole cluster and the parent of the other values in the same cluster. Despite this, the operations on red–black trees are more economical in time because you don't have to maintain the vector of values.^[20] It may be costly if values are stored directly in each node rather than being stored by reference. B-tree nodes, however, are more economical in space because you don't need to store the color attribute for each node. Instead, you have to know which slot in the cluster vector is used. If values are stored by reference, e. g. objects, null references can be used and so the cluster can be represented by a vector containing 3 slots for value pointers plus 4 slots for child references in the tree. In that case, the B-tree can be more compact in memory, improving data locality.

The same analogy can be made with B-trees with larger orders that can be structurally equivalent to a colored binary tree: you just need more colors. Suppose that you add blue, then the blue–red–black tree defined like red–black trees but with the additional constraint that no two successive nodes in the hierarchy will be blue and all blue nodes will be children of a red node, then it becomes equivalent to a B-tree whose clusters will have at most 7 values in the following colors: blue, red, blue, black, blue, red, blue (For each cluster, there will be at most 1 black node, 2 red nodes, and 4 blue nodes).

For moderate volumes of values, insertions and deletions in a colored binary tree are faster compared to B-trees because colored trees don't attempt to maximise the fill factor of each horizontal cluster of nodes (only the minimum fill factor is guaranteed in colored binary trees, limiting the number of splits or junctions of clusters). B-trees will be faster for performing rotations (because rotations will frequently occur within the same cluster rather than with multiple separate nodes in a colored binary tree). For storing large volumes, however, B-trees will be much faster as they will be more compact by grouping several children in the same cluster where they can be accessed locally.

All optimizations possible in B-trees to increase the average fill factors of clusters are possible in the equivalent multicolored binary tree. Notably, maximizing the average fill factor in a structurally equivalent B-tree is the same as reducing the total height of the multicolored tree, by

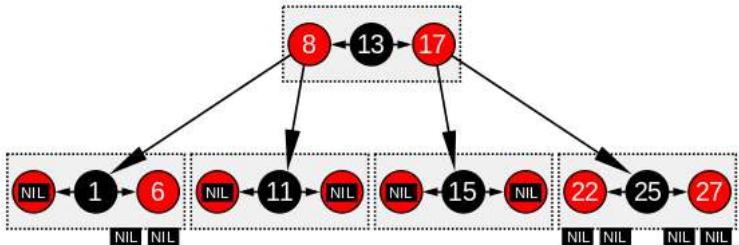


Figure 3: The same red–black tree as in the example above, now as a B-tree

increasing the number of non-black nodes. The worst case occurs when all nodes in a colored binary tree are black, the best case occurs when only a third of them are black (and the other two thirds are red nodes).

Applications and related data structures

Red–black trees offer worst-case guarantees for insertion time, deletion time, and search time. Not only does this make them valuable in time-sensitive applications such as real-time applications, but it makes them valuable building blocks in other data structures that provide worst-case guarantees; for example, many data structures used in computational geometry can be based on red–black trees, and the Completely Fair Scheduler used in current Linux kernels and epoll system call implementation^[21] uses red–black trees.

The AVL tree is another structure supporting $O(\log n)$ search, insertion, and removal. AVL trees can be colored red–black, thus are a subset of RB trees. Worst-case height is 0.720 times the worst-case height of RB trees, so AVL trees are more rigidly balanced. The performance measurements of Ben Pfaff with realistic test cases in 79 runs find AVL to RB ratios between 0.677 and 1.077, median at 0.947, and geometric mean 0.910.^[22] WAVL trees have a performance in between those two.

Red–black trees are also particularly valuable in functional programming, where they are one of the most common persistent data structures, used to construct associative arrays and sets that can retain previous versions after mutations. The persistent version of red–black trees requires $O(\log n)$ space for each insertion or deletion, in addition to time.

For every 2–4 tree, there are corresponding red–black trees with data elements in the same order. The insertion and deletion operations on 2–4 trees are also equivalent to color-flipping and rotations in red–black trees. This makes 2–4 trees an important tool for understanding the logic behind red–black trees, and this is why many introductory algorithm texts introduce 2–4 trees just before red–black trees, even though 2–4 trees are not often used in practice.

In 2008, Sedgewick introduced a simpler version of the red–black tree called the left-leaning red–black tree^[23] by eliminating a previously unspecified degree of freedom in the implementation. The LLRB maintains an additional invariant that all red links must lean left except during inserts and deletes. Red–black trees can be made isometric to either 2–3 trees,^[24] or 2–4 trees,^[23] for any sequence of operations. The 2–4 tree isometry was described in 1978 by Sedgewick.^[7] With 2–4 trees, the isometry is resolved by a "color flip," corresponding to a split, in which the red color of two children nodes leaves the children and moves to the parent node.

The original description of the tango tree, a type of tree optimised for fast searches, specifically uses red–black trees as part of its data structure.^[25]

As of Java 8, the HashMap has been modified such that instead of using a LinkedList to store different elements with colliding hashcodes, a red–black tree is used. This results in the improvement of time complexity of searching such an element from $O(m)$ to $O(\log m)$ where m is the number of elements with colliding hashcodes.^[26]

Operations

The read-only operations, such as search or tree traversal, on a red–black tree require no modification from those used for binary search trees, because every red–black tree is a special case of a simple binary search tree. However, the immediate result of an insertion or removal may violate the properties of a red–black tree, the restoration of which is called *rebalancing* so that

red–black trees become self-balancing. It requires in the *worst case* a small number, $O(\log n)$ in Big O notation, where n is the number of objects in the tree, on average or *amortized* $O(1)$, a constant number,^{[27]:310 [17]:158} of color changes (which are very quick in practice); and no more than three tree rotations^[28] (two for insertion).

If the example implementation below is not suitable, other implementations with explanations may be found in Ben Pfaff's^[29] annotated C library GNU libavl (<http://adtinfo.org/>) (v2.0.3 as of June 2019).

The details of the insert and removal operations will be demonstrated with example C++ code, which uses the type definitions, macros below, and the helper function for rotations:

```
// Basic type definitions:

enum color_t { BLACK, RED };

struct RBnode {      // node of red-black tree
    RBnode* parent; // == NIL if root of the tree
    RBnode* child[2]; // == NIL if child is empty
    // The index is:
    //   LEFT := 0, if (key < parent->key)
    //   RIGHT := 1, if (key > parent->key)
    enum color_t color;
    int key;
};

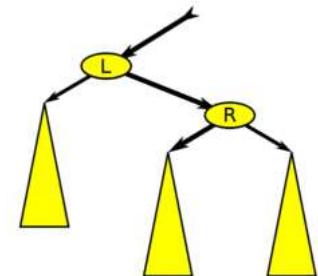
#define NIL NULL // null pointer or pointer to sentinel node
#define LEFT 0
#define RIGHT 1
#define left child[LEFT]
#define right child[RIGHT]

struct RBtree { // red-black tree
    RBnode* root; // == NIL if tree is empty
};

// Get the child direction (= { LEFT, RIGHT })
// of the non-root non-NIL RBnode* N:
#define childDir(N) (N == (N->parent)->right ? RIGHT : LEFT)
```

```
RBnode* RotateDirRoot(
    RBtree* T, // red-black tree
    RBnode* P, // root of subtree (may be the root of T)
    int dir) { // dir ∈ { LEFT, RIGHT }
    RBnode* G = P->parent;
    RBnode* S = P->child[1-dir];
    RBnode* C;
    assert(S != NIL); // pointer to true node required
    C = S->child[dir];
    P->child[1-dir] = C; if (C != NIL) C->parent = P;
    S->child[ dir] = P; P->parent = S;
    S->parent = G;
    if (G != NULL)
        G->child[ P == G->right ? RIGHT : LEFT ] = S;
    else
        T->root = S;
    return S; // new root of subtree
}

#define RotateDir(N,dir) RotateDirRoot(T,N,dir)
#define RotateLeft(N) RotateDirRoot(T,N,LEFT)
#define RotateRight(N) RotateDirRoot(T,N,RIGHT)
```



Left rotation and
right rotation,
animated.

Notes to the sample code and diagrams of insertion and removal

The proposal breaks down both, insertion and removal (not mentioning some very simple cases), into six constellations of nodes, edges and colors, which are called cases. The proposal contains for both, insertion and removal, exactly one case that advances one black level closer to the root and loops, the other five cases rebalance the tree of their own. The more complicated cases are pictured in a diagram.

-  symbolises a red node and  a (non-NIL) black node (of black height ≥ 1),  symbolises the color red or black of a non-NIL node, but the same color throughout the same diagram. NIL nodes are not represented in the diagrams.
 - The variable **N** denotes the current node, which is labeled  or  in the diagrams.
 - A diagram contains three columns and two to four actions. The left column shows the first iteration, the right column the higher iterations, the middle column shows the segmentation of a case into its different actions.^[30]
1. The action "entry" shows the constellation of nodes with their colors which defines a case and mostly violates some of the requirements.
A blue border rings the current node **N** and the other nodes are labeled according to their relation to **N**.
 2. If a rotation is considered useful, this is pictured in the next action, which is labeled "rotation".
 3. If some recoloring is considered useful, this is pictured in the next action, which is labeled "color".^[31]
 4. If there is still some need to repair, the cases make use of code of other cases and this after a reassignment of the current node **N**, which then again carries a blue ring and relative to which other nodes may have to be reassigned also. This action is labeled "reassign".
For both, insert and delete, there is (exactly) one case which iterates one black level closer to the root; then the reassigned constellation satisfies the respective loop invariant.
- A possibly numbered triangle with a black circle atop  represents a red-black subtree (connected to its parent according to requirement 3) with a black height equal to the iteration level minus one, i. e. zero in the first iteration. Its root may be red or black.
A possibly numbered triangle  represents a red-black subtree with a black height one less, i. e. its parent has black height zero in the second iteration.

Remark

For simplicity, the sample code uses the disjunction:

```
U == NIL || U->color == BLACK // considered black
```

and the conjunction:

```
U != NIL && U->color == RED // not considered black
```

Thereby, it must be kept in mind that both statements are *not* evaluated in total, if $U == \text{NIL}$. Then in both cases $U->\text{color}$ is not touched (see [Short-circuit evaluation](#)).
 (The comment considered black is in accordance with [requirement 2](#).)
 The related if-statements have to occur far less frequently if the proposal[\[30\]](#) is realised.

Insertion

Insertion begins by placing the new (non-NIL) node, say **N**, at the position in the [binary search tree](#) of a NIL node whose [in-order predecessor's](#) key compares less than the new node's key, which in turn compares less than the key of its in-order successor. (Frequently, this positioning is the result of a search within the tree immediately preceding the insert operation and consists of a node **P** together with a direction **dir** with $P->\text{child}[\text{dir}] == \text{NIL}$.) The newly inserted node is temporarily colored red so that all paths contain the same number of black nodes as before. But if its parent, say **P**, is also red then this action introduces a [red-violation](#).

```
void RBinsert1(
    RBtree* T,           // -> red-black tree
    struct RBnode* N,    // -> node to be inserted
    struct RBnode* P,    // -> parent node of N ( may be NULL )
    byte dir)            // side ( LEFT or RIGHT ) of P where to insert N
{
    struct RBnode* G;   // -> parent node of P
    struct RBnode* U;   // -> uncle of N

    N->color = RED;
    N->left = NIL;
    N->right = NIL;
    N->parent = P;
    if (P == NULL) {    // There is no parent
        T->root = N;   // N is the new root of the tree T.
        return; // insertion complete
    }
    P->child[dir] = N; // insert N as dir-child of P
    // start of the (do while)-loop:
    do {
```

The rebalancing loop of the insert operation has the following [invariant](#):

- The variable **N**, representing the current node **N** and initially the insertion node, is made the variable running through the loop.
- **N** is  (red) at the beginning of each iteration.
- [Requirement 3](#) is satisfied for all pairs node←parent with the possible exception **N←P** when **P** is also red (a [red-violation](#) at **N**).
- All other properties (including [requirement 4](#)) are satisfied throughout the tree.

Notes to the insert diagrams

- In the diagrams, **P** is used for **N**'s parent, **G** for its grandparent, and **U** for its uncle. In the table a — sign indicates root.
- The diagrams show the parent node **P** as the left child of its parent **G** even though it is possible for **P** to be on either side. The sample code covers both possibilities by means of the side variable **dir**.

- The diagrams show the cases where **P** is red also, the red-violation.
- The column *x* indicates the change in child direction, i.e. o (for "outer") means that **P** and **N** are both left or both right children, whereas i (for "inner") means that the child direction changes from **P**'s to **N**'s.

before				case	rotation	assignment	after				Δh	
P	G	U	x				P	G	U	x		
●				I1							✓	
●	●	●		I2		N:=G	?			?	2	
—				I3							✓	
●	—			I4			●				✓	
●	●	●	i	I5	P~N	N:=P	●	●	●	o	I6	0
●	●	●	o	I6	P~G		●	●	●		✓	

Insertion: This synopsis shows in its *before* columns, that all possible cases^[32] of constellations are covered.

- The column group *before* defines the case, whose name is given in the column *case*. Thereby possible values in cells left empty are ignored. So in case **I2** the sample code covers both possibilities of child directions of **N**, although the corresponding diagram shows only one.
- The rows in the synopsis are ordered such that the coverage of all possible RB cases is easily comprehensible.
- The column *rotation* indicates whether a rotation contributes to the rebalancing.
- The column *assignment* shows an assignment of **N** before entering a subsequent step. This possibly induces a reassignment of the other nodes **P**, **G**, **U** also.
- If something has been changed by the case, this is shown in the column group *after*.
- A ✓ sign in column *next* signifies that the rebalancing is complete with this step. If the column *after* determines exactly one case, this case is given as the subsequent one, otherwise there are question marks.
- The loop is contained in the sections "Insert case I1" and "Insert case I2", where in case **I2** the problem of rebalancing is escalated $\Delta h = 2$ tree levels or 1 black level higher in the tree, in that the grandfather **G** becomes the new current node **N**. So it takes maximally $\frac{h}{2}$ steps of iteration to repair the tree (where **h** is the height of the tree). Because the probability of escalation decreases exponentially with each step the total rebalancing cost is constant on average, indeed amortized constant.
- From the body of the loop, case **I1** exits by itself and there are exiting branches to cases **I4**, **I6**, **I5 + I6**, and **I3**.
- Rotations occur in cases **I6** and **I5 + I6** – outside the loop. Therefore, at most two rotations occur in total.

Insert case I1

The current node's parent **P** is black, so requirement 3 holds. Requirement 4 holds also according to the loop invariant.

```

if (P->color == BLACK) {
    // Case_I1 (P black):
    return; // insertion complete
}
// From now on P is red.
if ((G = P->parent) == NULL)
    goto Case_I4; // P red and root
// else: P red and G!=NULL.
dir = childDir(P); // the side of parent G on which node P is located
U = G->child[1-dir]; // uncle
if (U == NIL || U->color == BLACK) // considered black
    goto Case_I56; // P red && U black

```

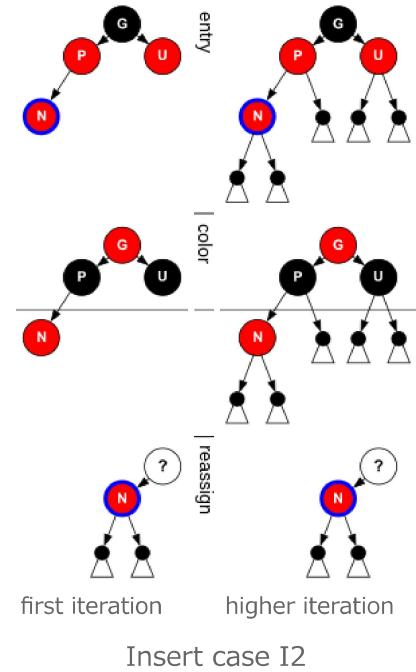
Insert case I2

If both the parent **P** and the uncle **U** are red, then both of them can be repainted black and the grandparent **G** becomes red for maintaining requirement 4. Since any path through the parent or uncle must pass through the grandparent, the number of black nodes on these paths has not changed. However, the grandparent **G** may now violate requirement 3, if it has a red parent. After relabeling **G** to **N** the loop invariant is fulfilled so that the rebalancing can be iterated on one black level (= 2 tree levels) higher.

```

// Case_I2 (P+U red):
P->color = BLACK;
U->color = BLACK;
G->color = RED;
N = G; // new current node
// iterate 1 black level higher
// (= 2 tree levels)
} while ((P = N->parent) != NULL);
// end of the (do while)-loop

```



Insert case I3

Insert case I2 has been executed for $\frac{h-1}{2}$ times and the total height of the tree has increased by 1, now being h . The current node **N** is the (red) root of the tree, and all RB-properties are satisfied.

```

// Leaving the (do while)-loop (after having fallen through from Case_I2).
// Case_I3: N is the root and red.
return; // insertion complete

```

Insert case I4

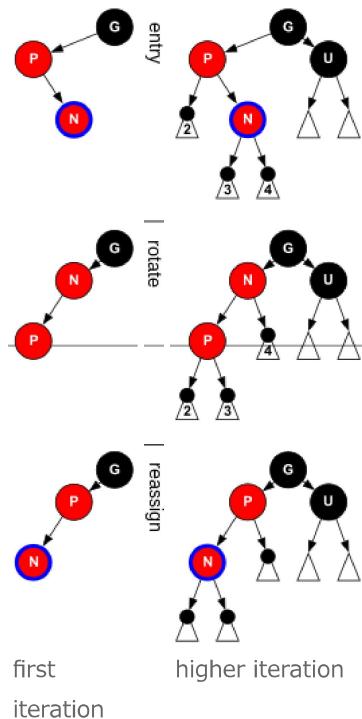
The parent **P** is red and the root. Because **N** is also red, requirement 3 is violated. But after switching **P**'s color the tree is in RB-shape. The black height of the tree increases by 1.

```

Case_I4: // P is the root and red:
P->color = BLACK;
return; // insertion complete

```

Insert case I5



Insert case I5

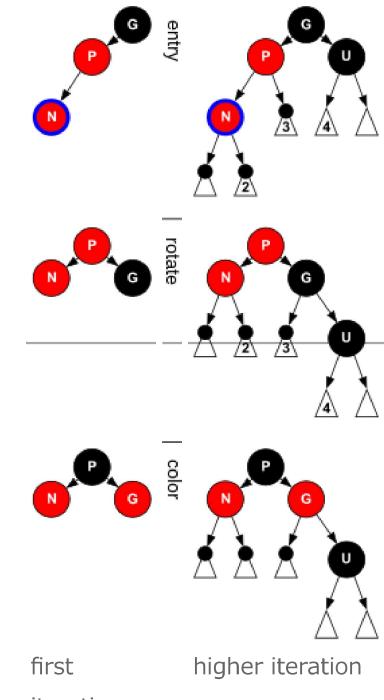
```
Case_I56: // P red && U black:
if (N == P->child[1-dir])
{ // Case_I5 (P red && U black && N inner grandchild of G):
    RotateDir(P,dir); // P is never the root
    N = P; // new current node
    P = G->child[dir]; // new parent of N
    // fall through to Case_I6
}
```

Insert case I6

The current node **N** is now certain to be an "outer" grandchild of **G** (left of left child or right of right child). Now (1-dir)-rotate at **G**, putting **P** in place of **G** and making **P** the parent of **N** and **G**. **G** is black and its former child **P** is red, since requirement 3 was violated. After switching the colors of **P** and **G** the resulting tree satisfies requirement 3. Requirement 4 also remains satisfied, since all paths that went through the black **G** now go through the black **P**.

```
// Case_I6 (P red && U black && N outer grandchild of G):
RotateDirRoot(T,G,1-dir); // G may be the root
P->color = BLACK;
G->color = RED;
return; // insertion complete
} // end of RBinsert1
```

Because the algorithm transforms the input without using an auxiliary data structure and using only a small amount of extra storage space for auxiliary variables it is in-place.



Insert case I6

Removal

Simple cases

The label **N** denotes the current node that at entry is the node to be deleted.

If **N** is the root that does not have a non-NIL child, it is replaced by a NIL node, after which the tree is empty—and in RB-shape.

If **N** has exactly one non-NIL child, it must be a red child, according to conclusion 5.

If **N** is a red node, it cannot have exactly one non-NIL child, because this would have to be black by requirement 3. Furthermore, it cannot have exactly one black child according to conclusion 5. As a consequence, the red node **N** is without any child and can simply be removed.

If **N** is a black node, it may have two red children, a single red child or no non-NIL child at all. If **N** has a single red child, it is simply replaced with this child after painting the latter black.

If **N** has two non-NIL children, an additional navigation to the minimum element in its right subtree (which is **N**'s in-order successor, say **y**) finds a node with no other node between **N** and **y** (as shown here). This node **y** does not have a left child and thus has at most one non-NIL child. If **y** is to be removed in **N**'s place, the red–black tree data related with **N** and **y**, i. e. the color of and the pointers to and from the two nodes, have to be exchanged. (As a result, the modified red–black tree is the same as before, except that the order between **N** and **y** is reversed.) This choice may result in one of the more simple cases above, but if **y** is without child and black we arrive at ...

Removal of a black non-root leaf

The complex case is when **N** is not the root, colored black and has no proper child (\Leftrightarrow only NIL children). In the first iteration, **N** is replaced by NIL.

```
void RBdelete2(
    RBtree* T,           // -> red-black tree
    struct RBnode* N)   // -> node to be deleted
{
    struct RBnode* P = N->parent; // -> parent node of N
    byte dir;           // side of P on which N is located (= { LEFT, RIGHT })
    struct RBnode* S;  // -> sibling of N
    struct RBnode* C;  // -> close nephew
    struct RBnode* D;  // -> distant nephew

    // P != NULL, since N is not the root.
    dir = childDir(N); // side of parent P on which the node N is located
    // Replace N at its parent P by NIL:
    P->child[dir] = NIL;
    goto Start_D;      // jump into the loop

    // start of the (do while)-loop:
    do {
        dir = childDir(N); // side of parent P on which node N is located
    Start_D:
        S = P->child[1-dir]; // sibling of N (has black height >= 1)
        D = S->child[1-dir]; // distant nephew
        C = S->child[ dir]; // close nephew
        if (S->color == RED)
            goto Case_D3;           // S red ==> P+C+D black
        // S is black:
        if (D != NIL && D->color == RED) // not considered black
            goto Case_D6;           // D red && S black
        if (C != NIL && C->color == RED) // not considered black
            goto Case_D5;           // C red && S+D black
        // Here both nephews are == NIL (first iteration) or black (later).
        if (P->color == RED)
            goto Case_D4;           // P red && C+S+D black
```

The rebalancing loop of the delete operation has the following invariant:

- At the beginning of each iteration the black height of **N** equals the iteration number minus one, which means that in the first iteration it is zero and that **N** is a true black node **●** in higher iterations.
- The number of black nodes on the paths through **N** is one less than before the deletion, whereas it is unchanged on all other paths, so that there is a black-violation

at **P** if other paths exist.

- All other properties (including requirement 3) are satisfied throughout the tree.

Notes to the delete diagrams

- In the diagrams below, **P** is used for **N**'s parent, **S** for the sibling of **N**, **C** (meaning *close nephew*) for **S**'s child in the same direction as **N**, and **D** (meaning *distant nephew*) for **S**'s other child (**S** cannot be a NIL node in the first iteration, because it must have black height one, which was the black height of **N** before its deletion, but **C** and **D** may be NIL nodes).

before				case	rotation	assignment	after				Δh
P	C	S	D				P	C	S	D	
—				D1							✓
●	●	●	●	D2		N:=P	?			?	1
●	●	●	●	D3	P \cap S	N:=N	●	●	●	●	D6 0
●	●	●	●	D4			●	●	●	●	D5 0
●	●	●	●	D5	C \cap S	N:=N	●	●	●	●	D4 0
●	●	●	●	D6	P \cap S		●	●	●	●	✓

Deletion: This synopsis shows in its *before* columns, that all possible cases^[32] of color constellations are covered.

- The diagrams show the current node **N** as the left child of its parent **P** even though it is possible for **N** to be on either side. The code samples cover both possibilities by means of the side variable *dir*.
- At the beginning (in the first iteration) of removal, **N** is the NIL node replacing the node to be deleted. Because its location in parent's node is the only thing of importance, it is symbolised by ~~✓~~ (meaning: the current node **N** is a NIL node and left child) in the left column of the delete diagrams. As the operation proceeds also proper nodes (of black height ≥ 1) may become current (see e. g. case D2).
- By counting the black bullets (● and ~~✓~~) in a delete diagram it can be observed that the paths through **N** have one bullet less than the other paths. This means a black-violation at **P**—if it exists.
- The color constellation in column group *before* defines the case, whose name is given in the column *case*. Thereby possible values in cells left empty are ignored.
- The rows in the synopsis are ordered such that the coverage of all possible RB cases is easily comprehensible.
- The column *rotation* indicates whether a rotation contributes to the rebalancing.
- The column *assignment* shows an assignment of **N** before entering a subsequent iteration step. This possibly induces a reassignment of the other nodes **P**, **C**, **S**, **D** also.
- If something has been changed by the case, this is shown in the column group *after*.
- A ✓ sign in column *next* signifies that the rebalancing is complete with this step. If the column *after* determines exactly one case, this case is given as the subsequent

one, otherwise there are question marks.

- The loop is contained in the sections from Start_D through "Delete case D2", where the problem of rebalancing is escalated $\Delta h = 1$ level higher in the tree in that the parent **P** becomes the new current node **N**. So it takes maximally **h** iterations to repair the tree (where **h** is the height of the tree). Because the probability of escalation decreases exponentially with each iteration the total rebalancing cost is constant on average, indeed amortized constant. (Just as an aside: Mehlhorn & Sanders point out: "AVL trees do *not* support constant amortized update costs."^{[17]:165,158} This is true for the rebalancing after a deletion, but not AVL insertion.^[33])
- Out of the body of the loop there are exiting branches to the cases **D3**, **D6**, **D5 + D6**, **D4**, and **D1**; section "Delete case D3" of its own has three different exiting branches to the cases **D6**, **D5** and **D4**.
- Rotations occur in cases **D6** and **D5 + D6** and **D3 + D5 + D6** – all outside the loop. Therefore, at most three rotations occur in total.

Delete case D1

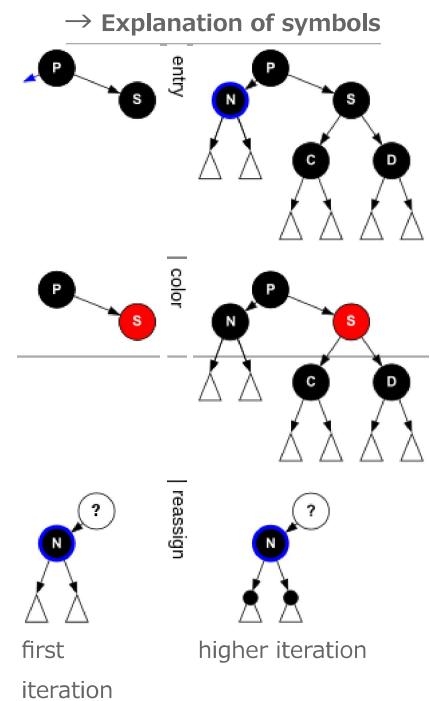
The current node **N** is the new root. One black node has been removed from every path, so the RB-properties are preserved. The black height of the tree decreases by 1.

```
// Case_D1 (P == NULL):
return; // deletion complete
```

Delete case D2

P, **S**, and **S**'s children are black. After painting **S** red all paths passing through **S**, which are precisely those paths *not* passing through **N**, have one less black node. Now all paths in the subtree rooted by **P** have the same number of black nodes, but one fewer than the paths that do not pass through **P**, so requirement 4 may still be violated. After relabeling **P** to **N** the loop invariant is fulfilled so that the rebalancing can be iterated on one black level (= 1 tree level) higher.

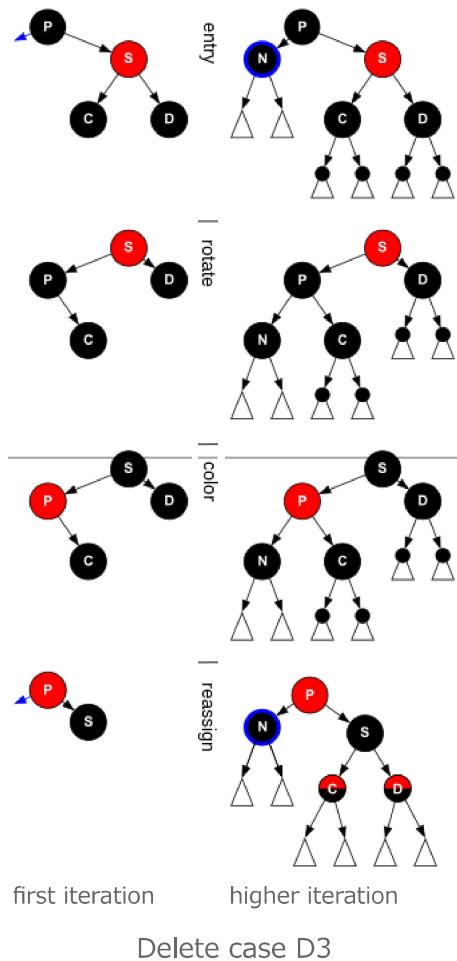
```
// Case_D2 (P+C+S+D black):
S->color = RED;
N = P; // new current node (maybe the root)
// iterate 1 black level
// (= 1 tree level) higher
} while ((P = N->parent) != NULL);
// end of the (do while)-loop
```



Delete case D3

The sibling **S** is red, so **P** and the nephews **C** and **D** have to be black. A **dir**-rotation at **P** turns **S** into **N**'s grandparent. Then after reversing the colors of **P** and **S**, the path through **N** is still

Delete case D2



short one black node. But **N** now has a red parent **P** and after the reassignment a black sibling **S**, so the transformations in cases D4, D5, or D6 are able to restore the RB-shape.

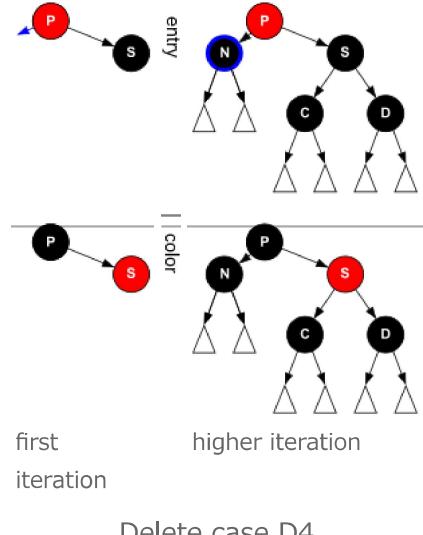
```
Case_D3: // S red && P+C+D black:
    RotateDirRoot(T, P, dir); // P may be the root
    P->color = RED;
    S->color = BLACK;
    S = C; // != NIL
    // now: P red && S black
    D = S->child[1-dir]; // distant nephew
    if (D != NIL && D->color == RED)
        goto Case_D6; // D red && S black
    C = S->child[ dir]; // close nephew
    if (C != NIL && C->color == RED)
        goto Case_D5; // C red && S+D black
    // Otherwise C+D considered black.
    // fall through to Case_D4
```

Delete case D4

The sibling **S** and **S**'s children are black, but **P** is red. Exchanging the colors of **S** and **P** does not affect the number of black nodes on paths going through **S**, but it does add one to the number of black nodes on paths going through **N**,

making up for the deleted black node on those paths.

```
Case_D4: // P red && S+C+D black:
    S->color = RED;
    P->color = BLACK;
    return; // deletion complete
```

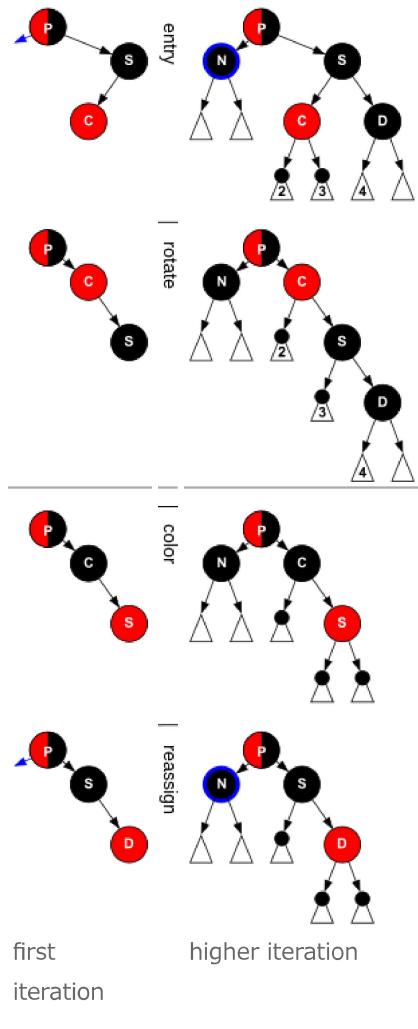


Delete case D5

The sibling **S** is black, **S**'s close child **C** is red, and **S**'s distant child **D** is black. After a (1-dir)-rotation at **S** the nephew **C** becomes **S**'s parent and **N**'s new sibling. The colors of **S** and **C** are exchanged. All paths still have the same number of black nodes, but now **N** has a black sibling whose distant child is red, so the constellation is fit for case D6. Neither **N** nor its parent **P** are affected by this transformation, and **P** may be red or black (● in the diagram).

```
Case_D5: // C red && S+D black:
    RotateDir(S, 1-dir); // S is never the root
    S->color = RED;
    C->color = BLACK;
    D = S;
    S = C;
    // now: D red && S black
    // fall through to Case_D6
```

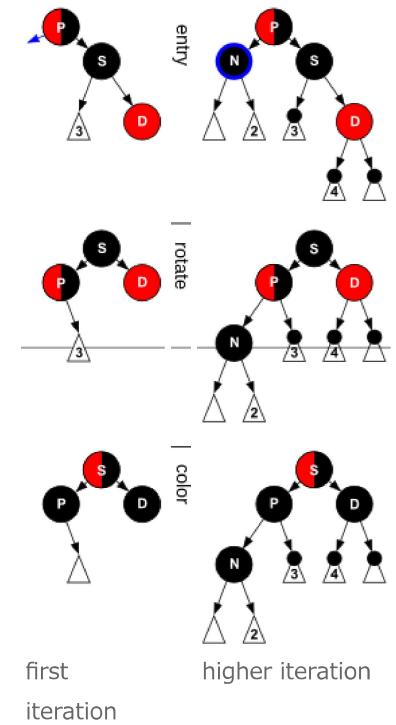
Delete case D6



Delete case D5

The sibling **S** is black, **S**'s distant child **D** is red. After a **dir**-rotation at **P** the sibling **S** becomes the parent of **P** and **S**'s distant child **D**. The colors of **P** and **S** are exchanged, and **D** is made black. The whole subtree still has the same color at its root **S**, namely either red or black (● in the diagram), which refers to the same color both before and after the transformation. This way requirement 3 is preserved. The paths in the subtree not passing through **N** (i.o.w. passing through **D** and node 3 in the diagram) pass through the same number of black nodes as before, but **N** now has one additional black ancestor: either **P** has become black, or it was black and **S** was added as a black grandparent. Thus, the paths passing through **N** pass through one additional black node, so that requirement 4 is restored and the total tree is in RB-shape.

```
Case_D6: // D red && S black:
    RotateDirRoot(T, P, dir); // P may be the root
    S->color = P->color;
    P->color = BLACK;
    D->color = BLACK;
    return; // deletion complete
} // end of RBdelete2
```



Delete case D6

Because the algorithm transforms the input without using an auxiliary data structure and using only a small amount of extra storage space for auxiliary variables it is in-place.

Proof of bounds

For $h \in \mathbb{N}$ there is a red-black tree of height h with

$$m_h = 2^{\lfloor (h+1)/2 \rfloor} + 2^{\lfloor h/2 \rfloor} - 2$$

$$= \begin{cases} 2 \cdot 2^{\frac{h}{2}} - 2 = 2^{\frac{h+1}{2}} - 2 & \text{if } h \text{ even} \\ 3 \cdot 2^{\frac{h-1}{2}} - 2 & \text{if } h \text{ odd} \end{cases}$$

nodes ($\lfloor \cdot \rfloor$ is the floor function) and there is no red-black tree of this tree height with fewer nodes—therefore it is **minimal**.

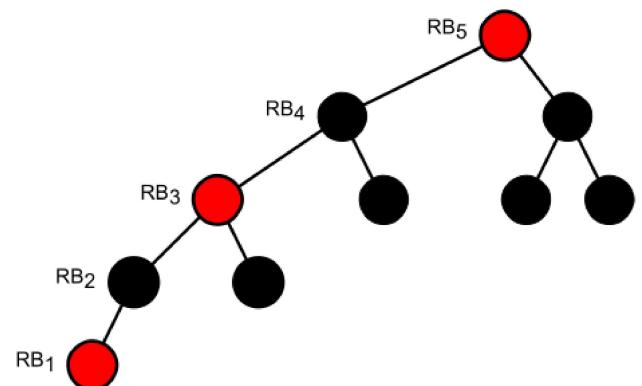


Figure 4: Red-black trees RB_h of heights $h=1,2,3,4,5$, each with minimal number 1,2,4,6 resp. 10 of nodes.

Its black height is $\lceil h/2 \rceil$ (with black root) or for odd h (then with a red root) also $(h-1)/2$.

Proof

For a red–black tree of a certain height to have minimal number of nodes, it must have exactly one longest path with maximal number of red nodes, to achieve a maximal tree height with a minimal black height. Besides this path all other nodes have to be black.^{[16]:444} Proof sketch If a node is taken off this tree it either loses height or some RB property.

The RB tree of height $h = 1$ with red root is minimal. This is in agreement with

$$m_1 = 2^{\lfloor(1+1)/2\rfloor} + 2^{\lfloor 1/2 \rfloor} - 2 = 2^1 + 2^0 - 2 = 1.$$

A minimal RB tree (RB_h in figure 4) of height $h > 1$ has a root whose two child subtrees are of different height. The higher child subtree is also a minimal RB tree, RB_{h-1} , containing also a longest path that defines its height $h-1$; it has m_{h-1} nodes and the black height $\lceil(h-1)/2\rceil =: s$. The other subtree is a perfect binary tree of (black) height s having $2^s - 1 = 2^{\lfloor(h-1)/2\rfloor} - 1$ black nodes—and no red node. Then the number of nodes is by induction

$$\begin{array}{llll} m_h = (m_{h-1}) & + & (1) & + (2^{\lfloor(h-1)/2\rfloor} - 1) \\ (\text{higher subtree}) & & (\text{root}) & (\text{second subtree}) \end{array}$$

resulting in

$$\begin{aligned} m_h &= (2^{\lfloor h/2 \rfloor} + 2^{\lfloor(h-1)/2\rfloor} - 2) + 2^{\lfloor(h-1)/2\rfloor} \\ &= 2^{\lfloor h/2 \rfloor} + 2^{\lfloor(h+1)/2\rfloor} - 2 \blacksquare \end{aligned}$$

The graph of the function m_h is convex and piecewise linear with breakpoints at $(h = 2k \mid m_{2k} = 2 \cdot 2^k - 2)$ where $k \in \mathbb{N}$. The function has been tabulated as $m_h = \text{A027383}(h-1)$ for $h \geq 1$ (sequence A027383 in the OEIS).

Solving the function for h

The inequality $9 > 8 = 2^3$ leads to $3 > 2^{3/2}$, which for odd h leads to

$$m_h = 3 \cdot 2^{(h-1)/2} - 2 = (3 \cdot 2^{-3/2}) \cdot 2^{(h+2)/2} - 2 > 2 \cdot 2^{h/2} - 2.$$

So in both, the even and the odd case, h is in the interval

$$\begin{array}{llll} \log_2(n+1) & \leq h \leq & 2\log_2(n+2) - 2 = 2\log_2(n/2 + 1) & [< 2\log_2(n+1)] \\ (\text{perfect} & & & \\ \text{binary} & & (\text{minimal red–black tree}) & \\ \text{tree}) & & & \end{array}$$

with n being the number of nodes.^[34]

Conclusion

A red–black tree with n nodes (keys) has tree height $h \in O(\log n)$.

Set operations and bulk operations

In addition to the single-element insert, delete and lookup operations, several set operations have been defined on red–black trees: union, intersection and set difference. Then fast *bulk* operations on insertions or deletions can be implemented based on these set functions. These set operations rely on two helper operations, *Split* and *Join*. With the new operations, the implementation of red–black trees can be more efficient and highly-parallelizable.^[35] In order to achieve its time complexities this implementation requires that the root is allowed to be either red or black, and that every node stores its own **black height**.

- *Join*: The function *Join* is on two red–black trees t_1 and t_2 and a key k , where $t_1 < k < t_2$, i. .e. all keys in t_1 are less than k , and all keys in t_2 are greater than k . It returns a tree containing all elements in t_1 , t_2 also as k .

If the two trees have the same black height, *Join* simply creates a new node with left subtree t_1 , root k and right subtree t_2 . If both t_1 and t_2 have black root, set k to be red. Otherwise k is set black.

If the black heights are unequal, suppose that t_1 has larger black height than t_2 (the other case is symmetric). *Join* follows the right spine of t_1 until a black node c , which is balanced with t_2 . At this point a new node with left child c , root k (set to be red) and right child t_2 is created to replace c . The new node may invalidate the red–black invariant because at most three red nodes can appear in a row. This can be fixed with a double rotation. If double red issue propagates to the root, the root is then set to be black, restoring the properties. The cost of this function is the difference of the black heights between the two input trees.

- *Split*: To split a red–black tree into two smaller trees, those smaller than key x , and those larger than key x , first draw a path from the root by inserting x into the red–black tree. After this insertion, all values less than x will be found on the left of the path, and all values greater than x will be found on the right. By applying *Join*, all the subtrees on the left side are merged bottom-up using keys on the path as intermediate nodes from bottom to top to form the left tree, and the right part is symmetric.

For some applications, *Split* also returns a boolean value denoting if x appears in the tree. The cost of *Split* is $O(\log n)$, order of the height of the tree. This algorithm actually has nothing to do with any special properties of a red–black tree, and may be used on any tree with a *join* operation, such as an AVL tree.

The join algorithm is as follows:

```
function joinRightRB(TL, k, TR):
    if (TL.color=black) and (TL.blackHeight=TR.blackHeight):
        return Node(TL, (k, red), TR)
    T'=Node(TL.left, (TL.key, TL.color), joinRightRB(TL.right, k, TR))
    if (TL.color=black) and (T'.right.color=T'.right.right.color=red):
```

```

T'.right.right.color=black;
return rotateLeft(T')
return T' /* T' [recte T'] */

function joinLeftRB(TL, k, TR):
/* symmetric to joinRightRB */

function join(TL, k, TR):
if TL.blackHeight>TR.blackHeight:
T'=joinRightRB(TL, k, TR)
if (T'.color=red) and (T'.right.color=red):
T'.color=black
return T'
if TR.blackHeight>TL.blackHeight:
/* symmetric */
if (TL.color=black) and (TR.color=black):
return Node(TL,(k,red),TR)
return Node(TL,(k,black),TR)

```

The split algorithm is as follows:

```

function split(T, k):
if (T = nil) return (nil, false, nil)
if (k = T.key) return (T.left, true, T.right)
if (k < T.key):
(L', b, R') = split(T.left, k)
return (L', b, join(R', T.key, T.right))
(L', b, R') = split(T.right, k)
return (join(T.left, T.key, L'), b, T.right)

```

The union of two red–black trees t_1 and t_2 representing sets A and B , is a red–black tree t that represents $A \cup B$. The following recursive function computes this union:

```

function union(t1, t2):
if t1 = nil return t2
if t2 = nil return t1
(L1, b, R1)=split(t1, t2.key)
proc1=start:
    TL=union(L1, t2.left)
proc2=start:
    TR=union(R1, t2.right)
wait all proc1,proc2
return join(TL, t2.key, TR)

```

Here, *split* is presumed to return two trees: one holding the keys less its input key, one holding the greater keys. (The algorithm is non-destructive, but an in-place destructive version exists also.)

The algorithm for intersection or difference is similar, but requires the *Join2* helper routine that is the same as *Join* but without the middle key. Based on the new functions for union, intersection or difference, either one key or multiple keys can be inserted to or deleted from the red–black tree. Since *Split* calls *Join* but does not deal with the balancing criteria of red–black trees directly, such an implementation is usually called the "join-based" implementation.

The complexity of each of union, intersection and difference is $O\left(m \log\left(\frac{n}{m} + 1\right)\right)$ for two red–black trees of sizes m and $n (\geq m)$. This complexity is optimal in terms of the number of comparisons. More importantly, since the recursive calls to union, intersection or difference are independent of each other, they can be executed in parallel with a parallel depth $O(\log m \log n)$

[\[35\]](#) When $m = 1$, the join-based implementation has the same computational directed acyclic graph (DAG) as single-element insertion and deletion if the root of the larger tree is used to split the smaller tree.

Parallel algorithms

Parallel algorithms for constructing red–black trees from sorted lists of items can run in constant time or $O(\log \log n)$ time, depending on the computer model, if the number of processors available is asymptotically proportional to the number n of items where $n \rightarrow \infty$. Fast search, insertion, and deletion parallel algorithms are also known.[\[36\]](#)

The join-based algorithms for red–black trees are parallel for bulk operations, including union, intersection, construction, filter, map-reduce, and so on.

Parallel bulk operations

Basic operations like insertion, removal or update can be parallelised by defining operations that process bulks of multiple elements. It is also possible to process bulks with several basic operations, for example bulks may contain elements to insert and also elements to remove from the tree.

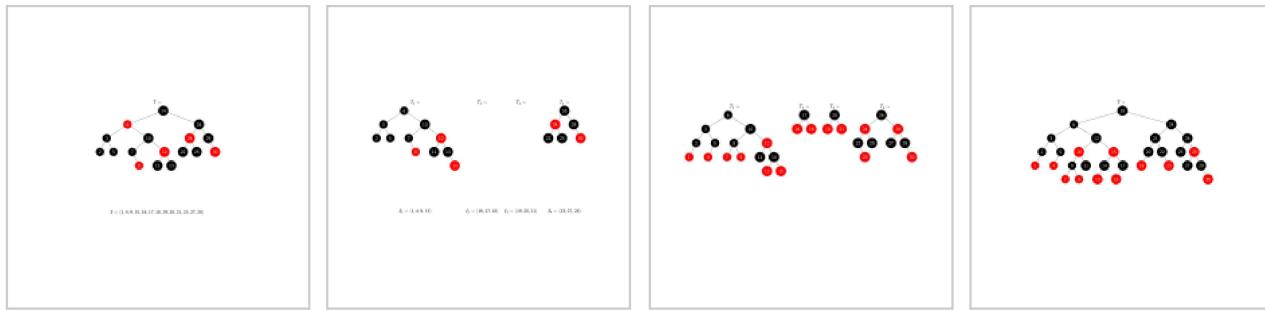
The algorithms for bulk operations aren't just applicable to the red–black tree, but can be adapted to other sorted sequence data structures also, like the [2–3 tree](#), [2–3–4 tree](#) and [\(a,b\)-tree](#). In the following different algorithms for bulk insert will be explained, but the same algorithms can also be applied to removal and update. Bulk insert is an operation that inserts each element of a sequence I into a tree T .

Join-based

This approach can be applied to every sorted sequence data structure that supports efficient join- and split-operations.[\[37\]](#) The general idea is to split I and T in multiple parts and perform the insertions on these parts in parallel.

1. First the bulk I of elements to insert must be sorted.
2. After that, the algorithm splits I into $k \in \mathbb{N}^+$ parts $\langle I_1, \dots, I_k \rangle$ of about equal sizes.
3. Next the tree T must be split into k parts $\langle T_1, \dots, T_k \rangle$ in a way, so that for every $j \in \mathbb{N}^+ | 1 \leq j < k$ following constraints hold:
 1. $\text{last}(I_j) < \text{first}(T_{j+1})$
 2. $\text{last}(T_j) < \text{first}(I_{j+1})$
4. Now the algorithm inserts each element of I_j into T_j sequentially. This step must be performed for every j , which can be done by up to k processors in parallel.
5. Finally, the resulting trees will be joined to form the final result of the entire operation.

Note that in Step 3 the constraints for splitting I assure that in Step 5 the trees can be joined again and the resulting sequence is sorted.



initial tree

split I and T

insert into the split
T

join T

The pseudo code shows a simple divide-and-conquer implementation of the join-based algorithm for bulk-insert. Both recursive calls can be executed in parallel. The join operation used here differs from the version explained in this article, instead join2 is used, which misses the second parameter k.

```

bulkInsert(T, I, k):
    I.sort()
    bulkInsertRec(T, I, k)

bulkInsertRec(T, I, k):
    if k = 1:
        forall e in I: T.insert(e)
    else
        m := [size(I) / 2]
        (T1, _, T2) := split(T, I[m])
        bulkInsertRec(T1, I[0 .. m], [k / 2])
        || bulkInsertRec(T2, I[m + 1 .. size(I) - 1], [k / 2])
        T ← join2(T1, T2)
    
```

Execution time

Sorting I is not considered in this analysis.

#recursion levels	$\in O(\log k)$
T(split) + T(join)	$\in O(\log T)$
insertions per thread	$\in O\left(\frac{ I }{k}\right)$
T(insert)	$\in O(\log T)$
T(bulkInsert) with $k = \#processors$	$\in O\left(\log k \log T + \frac{ I }{k} \log T \right)$

This can be improved by using parallel algorithms for splitting and joining. In this case the execution time is $\in O\left(\log |T| + \frac{|I|}{k} \log |T|\right)$.^[38]

Work

#splits, #joins	$\in O(k)$
W(split) + W(join)	$\in O(\log T)$
#insertions	$\in O(I)$
W(insert)	$\in O(\log T)$
W(bulkInsert)	$\in O(k \log T + I \log T)$

Pipelining

Another method of parallelizing bulk operations is to use a pipelining approach.^[39] This can be done by breaking the task of processing a basic operation up into a sequence of subtasks. For multiple basic operations the subtasks can be processed in parallel by assigning each subtask to a separate processor.

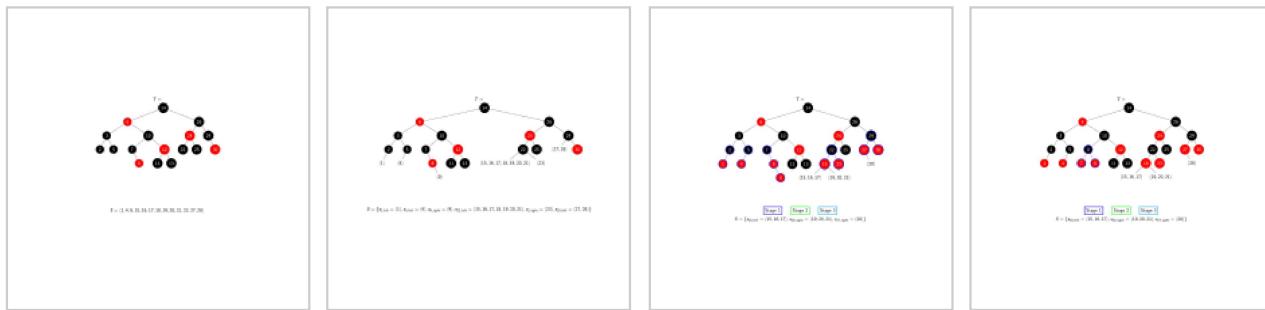
1. First the bulk I of elements to insert must be sorted.
2. For each element in I the algorithm locates the according insertion position in T . This can be done in parallel for each element $\in I$ since T won't be mutated in this process. Now I must be divided into subsequences S according to the insertion position of each element. For example $s_{n,\text{left}}$ is the subsequence of I that contains the elements whose insertion position would be to the left of node n .
3. The middle element $m_{n,\text{dir}}$ of every subsequence $s_{n,\text{dir}}$ will be inserted into T as a new node n' . This can be done in parallel for each $m_{n,\text{dir}}$ since by definition the insertion position of each $m_{n,\text{dir}}$ is unique. If $s_{n,\text{dir}}$ contains elements to the left or to the right of $m_{n,\text{dir}}$, those will be contained in a new set of subsequences S as $s_{n',\text{left}}$ or $s_{n',\text{right}}$.
4. Now T possibly contains up to two consecutive red nodes at the end of the paths from the root to the leaves, which needs to be repaired. Note that, while repairing, the insertion position of elements $\in S$ have to be updated, if the corresponding nodes are affected by rotations.

If two nodes have different nearest black ancestors, they can be repaired in parallel. Since at most four nodes can have the same nearest black ancestor, the nodes at the lowest level can be repaired in a constant number of parallel steps.

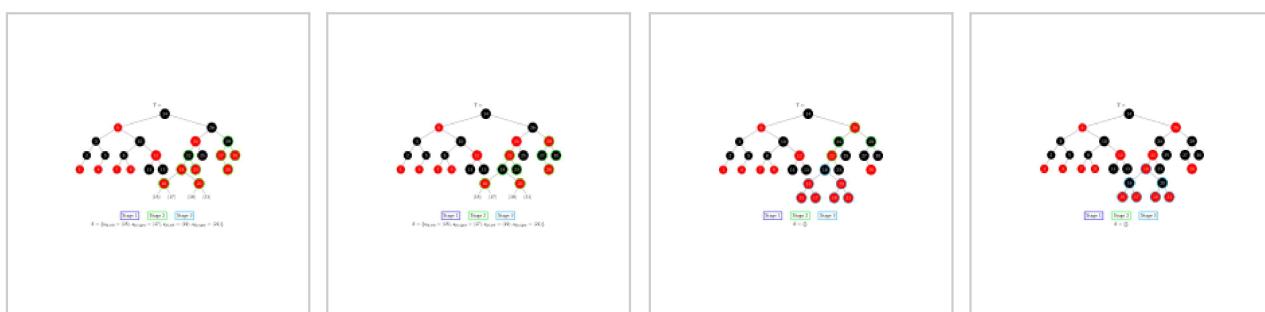
This step will be applied successively to the black levels above until T is fully repaired.

5. The steps 3 to 5 will be repeated on the new subsequences until S is empty. At this point every element $\in I$ has been inserted. Each application of these steps is called a *stage*. Since the length of the subsequences in S is $\in O(|I|)$ and in every stage the

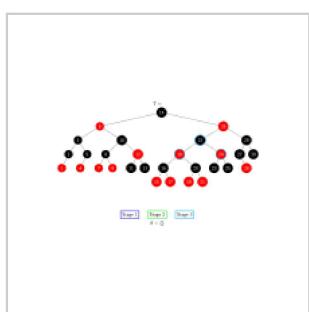
subsequences are being cut in half, the number of stages is $\in O(\log |I|)$. Since all stages move up the black levels of the tree, they can be parallelised in a pipeline. Once a stage has finished processing one black level, the next stage is able to move up and continue at that level.



Initial tree Find positions insert Stage 1 elements inserts Stage 1 begins to repair nodes



Stage 2 inserts elements Stage 2 begins to repair nodes Stage 3 inserts elements Stage 3 begins to repair nodes



Stage 3 continues to repair nodes

Execution time

Sorting I is not considered in this analysis. Also, $|I|$ is assumed to be smaller than $|T|$, otherwise it would be more efficient to construct the resulting tree from scratch.

T(find insert position)	$\in O(\log T)$
#stages	$\in O(\log I)$
T(insert) + T(repair)	$\in O(\log T)$
T(bulkInsert) with $I \sim \#processors$	$\in O(\log I + 2 \cdot \log T) = O(\log T)$

Work

W(find insert positions)	$\in O(I \log T)$
#insertions, #repairs	$\in O(I)$
W(insert) + W(repair)	$\in O(\log T)$
W(bulkInsert)	$\in O(2 \cdot I \log T) = O(I \log T)$

Popular culture

A red–black tree was referenced correctly in an episode of [Missing](#)^[40] as noted by [Robert Sedgewick](#) in one of his lectures:^[41]

Jess: "It was the red door again."

Pollock: "I thought the red door was the storage container."

Jess: "But it wasn't red anymore, it was black."

Antonio: "So red turning to black means what?"

Pollock: "Budget deficits, red ink, black ink."

Antonio: "It could be from a binary search tree. The red–black tree tracks every simple path from a node to a descendant leaf that has the same number of black nodes."

Jess: "Does that help you with the ladies?"

See also

- [List of data structures](#)
- [Tree data structure](#)
- [Tree rotation](#)
- [AA tree](#), a variation of the red–black tree
- [Left-leaning red–black tree](#)
- [AVL tree](#)
- [B-tree](#) ([2–3 tree](#), [2–3–4 tree](#), [B+ tree](#), [B*-tree](#), [UB-tree](#))
- [Scapegoat tree](#)
- [Splay tree](#)
- [T-tree](#)

■ WAVL tree

References and notes

1. Paton, James. "Red–Black Trees" (<http://pages.cs.wisc.edu/~paton/readings/Red-Black-Trees/>).
2. rebalancing only (no lookup), see Tarjan and Mehlhorn.
3. Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford (2001). "Red–Black Trees". *Introduction to Algorithms* (2nd ed.). MIT Press. pp. 273 ([https://archive.org/details/introductiontoal00corm_691/page/n735\)-301](https://archive.org/details/introductiontoal00corm_691/page/n735)-301). ISBN 978-0-262-03293-3.
4. Morris, John (1998). "Red–Black Trees" (https://www.cs.auckland.ac.nz/software/AlgAnim/red_black.html). *Data Structures and Algorithms*.
5. Bayer, Rudolf (1972). "Symmetric binary B-Trees: Data structure and maintenance algorithms". *Acta Informatica*. 1 (4): 290–306. doi:10.1007/BF00289509 (<https://doi.org/10.1007%2FBF00289509>). S2CID 28836825 (<https://api.semanticscholar.org/CorpusID:28836825>).
6. Drozdek, Adam (2001). *Data Structures and Algorithms in Java* (2 ed.). Sams Publishing. p. 323. ISBN 978-0534376680.
7. Guibas, Leonidas J.; Sedgewick, Robert (1978). "A Dichromatic Framework for Balanced Trees". *Proceedings of the 19th Annual Symposium on Foundations of Computer Science*. pp. 8–21. doi:10.1109/SFCS.1978.3 (<https://doi.org/10.1109%2FSFCS.1978.3>).
8. "Red Black Trees" (https://web.archive.org/web/20070927222509/http://eternallyconfuzzled.com/tuts/datastructures/jsw_tut_rbtree.aspx). *eternallyconfuzzled.com*. Archived from the original (http://eternallyconfuzzled.com/tuts/datastructures/jsw_tut_rbtree.aspx) on 2007-09-27. Retrieved 2015-09-02.
9. Sedgewick, Robert (2012). *Red–Black BSTs* (<https://www.coursera.org/learn/algorithms-graphs-data-structures/lecture/8acpe/red-black-trees>). Coursera. "A lot of people ask why did we use the name red–black. Well, we invented this data structure, this way of looking at balanced trees, at Xerox PARC that was the home of the personal computer and many other innovations that we live with today entering[sic] graphic user interfaces, Ethernet and object-oriented programmings[sic] and many other things. But one of the things that was invented there was laser printing and we were very excited to have nearby color laser printer that could print things out in color and out of the colors the red looked the best. So, that's why we picked the color red to distinguish red links, the types of links, in three nodes. So, that's an answer to the question for people that have been asking."
10. "Where does the term "Red/Black Tree" come from?" (<http://programmers.stackexchange.com/a/116621/40127>). *programmers.stackexchange.com*. Retrieved 2015-09-02.
11. Andersson, Arne (1993-08-11). "Balanced search trees made simple" (<https://web.archive.org/web/20181208183054/https://www.springer.com/la/book/9783540571551>). In Dehne, Frank; Sack, Jörg-Rüdiger; Santoro, Nicola; Whitesides, Sue (eds.). *Algorithms and Data Structures* (Proceedings). Lecture Notes in Computer Science. Vol. 709. Springer-Verlag Berlin Heidelberg. pp. 60–71. CiteSeerX 10.1.1.118.6192 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.118.6192>). doi:10.1007/3-540-57155-8_236 (https://doi.org/10.1007%2F3-540-57155-8_236). ISBN 978-3-540-57155-1. Archived from the original (<https://www.springer.com/la/book/9783540571551>) on 2018-12-08. Alt URL (<http://user.it.uu.se/~arnea/ps/simple.pdf>)

12. Okasaki, Chris (1999-01-01). "Red–black trees in a functional setting" (<https://doi.org/10.1017%2FS0956796899003494>). *Journal of Functional Programming*. 9 (4): 471–477. doi:10.1017/S0956796899003494 (<https://doi.org/10.1017%2FS0956796899003494>). ISSN 1469-7653 (<https://www.worldcat.org/issn/1469-7653>). S2CID 20298262 (<https://api.semanticscholar.org/CorpusID:20298262>).
13. Sedgewick, Robert (1983). *Algorithms* (<https://archive.org/details/algorithms00sedg>) (1st ed.). Addison-Wesley. ISBN 978-0-201-06672-2.
14. Sedgewick, Robert; Wayne, Kevin. "RedBlackBST.java" (<http://algs4.cs.princeton.edu/code/edu/princeton/cs/algs4/RedBlackBST.java.html>). *algs4.cs.princeton.edu*. Retrieved 7 April 2018.
15. Sedgewick, Robert (2008). "Left-leaning Red–Black Trees" (<https://sedgewick.io/talks/>).
16. Sedgewick, Robert; Wayne, Kevin (2011). *Algorithms* (<http://algs4.cs.princeton.edu>) (4th ed.). Addison-Wesley Professional. ISBN 978-0-321-57351-3.
17. Mehlhorn, Kurt; Sanders, Peter (2008). "7. Sorted Sequences" (<http://people.mpi-inf.mpg.de/~mehlhorn/ftp/Toolbox/SortedSequences.pdf>) (PDF). *Algorithms and Data Structures: The Basic Toolbox* (<http://people.mpi-inf.mpg.de/~mehlhorn/Toolbox.html>). Berlin/Heidelberg: Springer. CiteSeerX 10.1.1.148.2305 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.148.2305>). doi:10.1007/978-3-540-77978-0 (<https://doi.org/10.1007%2F978-3-540-77978-0>). ISBN 978-3-540-77977-3.
18. Cormen, Thomas; Leiserson, Charles; Rivest, Ronald; Stein, Clifford (2022). "13. Red–Black Trees". *Introduction to Algorithms* (4th ed.). MIT Press. pp. 331 (https://archive.org/details/introductiontoalgorithms_805/page/n328)–332. ISBN 9780262046305.
19. Using Knuth's definition of order: the maximum number of children
20. Sedgewick, Robert (1998). *Algorithms in C++* (<https://archive.org/details/algorithmscparts00sedg>) (1st ed.). Addison-Wesley Professional. pp. 565 (<https://archive.org/details/algorithmscparts00sedg>)–575. ISBN 978-0-201-35088-3.
21. "The Implementation of epoll (1)" (<https://idndx.com/2014/09/01/the-implementation-of-epoll-1/>). September 2014.
22. Pfaff 2004
23. "Robert Sedgewick" (<http://www.cs.princeton.edu/~rs/talks/LLRB/RedBlack.pdf>) (PDF). *Cs.princeton.edu*. 4 June 2020. Retrieved 26 March 2022.
24. "Balanced Trees" (<http://www.cs.princeton.edu/courses/archive/fall08/cos226/lectures/10BalancedTrees-2x2.pdf>) (PDF). *Cs.princeton.edu*. Retrieved 26 March 2022.
25. Demaine, E. D.; Harmon, D.; Iacono, J.; Pătrașcu, M. (2007). "Dynamic Optimality—Almost" (http://erikdemaine.org/papers/Tango_SICOMP/paper.pdf) (PDF). *SIAM Journal on Computing*. 37 (1): 240. doi:10.1137/S0097539705447347 (<https://doi.org/10.1137%2FS0097539705447347>). S2CID 1480961 (<https://api.semanticscholar.org/CorpusID:1480961>).
26. "How does a HashMap work in JAVA" (http://coding-geek.com/how-does-a-hashmap-work-in-java/#JAVA_8_improvements). *coding-geek.com*.
27. Tarjan, Robert Endre (April 1985). "Amortized Computational Complexity" (http://www.cs.duke.edu/courses/fall11/cps234/reading/Tarjan85_AmortizedComplexity.pdf) (PDF). *SIAM Journal on Algebraic and Discrete Methods*. 6 (2): 306–318. doi:10.1137/0606031 (<https://doi.org/10.1137%2F0606031>).
28. The important thing about these tree rotations is that they preserve the in-order sequence of the tree's nodes.
29. "Ben Pfaff (2007): Online HTML version of a well-documented collection of binary search tree and balanced tree library routines" (<https://mirrors.sunsite.dk/gnu/avl/>).

30. The left columns contain far less nodes than the right ones, especially for removal. This indicates that some efficiency can be gained by pulling the first iteration out of the rebalancing loops of insertion and deletion, because many of the named nodes are NIL nodes in the first iteration and definitively non-NIL later. (See also [this remark](#).)
31. Rotations have been placed before recoloring for reasons of clarity. But the two commute, so that it is free choice to move the rotation to the [tail](#).
32. The same partitioning is found in [Ben Pfaff](#).
33. Dinesh P. Mehta, Sartaj Sahni (Ed.) *Handbook of Data Structures and Applications* 10.4.2
34. Equality at the upper bound holds for the minimal RB trees RB_{2k} of even height $2 \cdot k$ with $n = 2 \cdot 2^k - 2$ nodes and only for those. So the inequality is marginally more precise than the widespread $h < 2 \log_2(n + 1)$, e. g. in [Cormen](#) p. 264. Moreover, these trees are binary trees that admit one [and only](#) one coloring conforming to the RB requirements 1 to 4. But there are further such trees, e. g. appending a child node to a black leaf always forces it to red. (A minimal RB tree of odd height allows to flip the root's color from red to black.)
35. Blelloch, Guy E.; Ferizovic, Daniel; Sun, Yihan (2016), "Just Join for Parallel Ordered Sets" (<http://www.cs.cmu.edu/~yihans/papers/join.pdf>) (PDF), *Symposium on Parallel Algorithms and Architectures, Proc. of 28th ACM Symp. Parallel Algorithms and Architectures (SPAA 2016)*, ACM, pp. 253–264, arXiv:1602.02120 (<https://arxiv.org/abs/1602.02120>), doi:10.1145/2935764.2935768 (<https://doi.org/10.1145%2F2935764.2935768>), ISBN 978-1-4503-4210-0, S2CID 2897793 (<https://api.semanticscholar.org/CorpusID:2897793>).
36. Park, Heejin; Park, Kunsoo (2001). "Parallel algorithms for red–black trees" (<https://doi.org/10.1016/S0304-3975%2800%2900287-5>). *Theoretical Computer Science*. 262 (1–2): 415–435. doi:10.1016/S0304-3975(00)00287-5 (<https://doi.org/10.1016/S0304-3975%2800%2900287-5>). "Our parallel algorithm for constructing a red–black tree from a sorted list of n items runs in $O(1)$ time with n processors on the CRCW PRAM and runs in $O(\log \log n)$ time with $n/\log \log n$ processors on the EREW PRAM."
37. Sanders, Peter (2019). Mehlhorn, Kurt; Dietzfelbinger, Martin; Dementiev, Roman (eds.). *Sequential and Parallel Algorithms and Data Structures : The Basic Toolbox*. Springer eBooks. Cham: Springer. pp. 252–253. doi:10.1007/978-3-030-25209-0 (<https://doi.org/10.1007%2F978-3-030-25209-0>). ISBN 9783030252090. S2CID 201692657 (<https://api.semanticscholar.org/CorpusID:201692657>).
38. Akhremtsev, Yaroslav; Sanders, Peter (2016). "Fast Parallel Operations on Search Trees". *HiPC 2016, the 23rd IEEE International Conference on High Performance Computing, Data, and Analytics, Hyderabad, India, December, 19–22*. IEEE, Piscataway (NJ): 291–300. arXiv:1510.05433 (<https://arxiv.org/abs/1510.05433>). Bibcode:2015arXiv151005433A (<https://ui.adsabs.harvard.edu/abs/2015arXiv151005433A>). ISBN 978-1-5090-5411-4.
39. Jájá, Joseph (1992). *An introduction to parallel algorithms* (<http://zbmath.org/?q=an:0781.68009>). Reading, Mass. [u.a.]: Addison-Wesley. pp. 65–70. ISBN 0201548569. Zbl 0781.68009 (<https://zbmath.org/?format=complete&q=an:0781.68009>).
40. *Missing (Canadian TV series)*. A, W Network (Canada); Lifetime (United States).

41. Robert Sedgewick (2012). *B-Trees* (<https://www.coursera.org/lecture/algorithms-part1/b-trees-optional-HIIHd>). Coursera. 9:48 minutes in. "So not only is there some excitement in that dialogue but it's also technically correct that you don't often find with math in popular culture of computer science. A red-black tree tracks every simple path from a node to a descendant leaf with the same number of black nodes they got that right."

Further reading

- Mathworld: Red-Black Tree (<http://mathworld.wolfram.com/Red-BlackTree.html>)
- San Diego State University: CS 660: Red-Black tree notes (<http://www.elis.sdsu.edu/courses/fall95/cs660/notes/RedBlackTree/RedBlack.html#RTFToC2>), by Roger Whitney
- Pfaff, Ben (June 2004). "Performance Analysis of BSTs in System Software" (<http://www.stanford.edu/~blp/papers/libavl.pdf>) (PDF). Stanford University.

External links

- Ben Pfaff: *An Introduction to Binary Search Trees and Balanced Trees*. Free Software Foundation, Boston 2004, [ftp.gnu.org \(https://ftp.gnu.org/gnu/avl/avl-2.0.3.pdf.gz\)](ftp://ftp.gnu.org/gnu/avl/avl-2.0.3.pdf.gz) (PDF gzip; 1662 kB)
 - A complete and working implementation in C ([https://web.archive.org/web/20140328232325/http://en.literateprograms.org/Red-black_tree_\(C\)](https://web.archive.org/web/20140328232325/http://en.literateprograms.org/Red-black_tree_(C)))
 - OCW MIT Lecture on Red-black Trees (<http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-046j-introduction-to-algorithms-sma-5503-fall-2005/video-lectures/lecture-10-red-black-trees-rotations-insertions-deletions/>) by Erik Demaine
 - Binary Search Tree Insertion Visualization (https://www.youtube.com/watch?v=_VbTnLV8plU) on YouTube – Visualization of random and pre-sorted data insertions, in elementary binary search trees, and left-leaning red-black trees
 - An intrusive red-black tree written in C++ (<https://gist.github.com/pallas/10697727>)
 - Red-black BSTs in 3.3 Balanced Search Trees (<https://algs4.cs.princeton.edu/33balanced/>)
 - Red-black BST Demo (<https://web.archive.org/web/20160722175219/http://algs4.cs.princeton.edu/lectures/33DemoRedBlackBST.mov>)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Red-black_tree&oldid=1178654552"

■