Drawing Trees Nicely with TEX*

Anne Brüggemann-Klein† Derick Wood‡

September 18, 2002

Abstract

Various algorithms have been proposed for the difficult problem of producing aesthetically pleasing drawings of trees, see [**?**, **?**] but implementations only exist as "special purpose software", designed for special environments. Therefore, many users resort to the drawing facilities available on most personal computers, but the figures obtained in this way still look "hand-drawn"; their quality is inferior to the quality of the surrounding text that can be realized by today's high quality text processing systems.

In this paper we present an entirely new solution that integrates a tree drawing algorithm into one of the best text processing systems available. More precisely, we present a TEX macro package TreeTEX that produces a drawing of a tree from a purely logical description. Our approach has three advantages. First, labels for nodes can be handled in a reasonable way. On the one hand, the tree drawing algorithm can compute the widths of the labels and take them into account for the positioning of the nodes; on the other hand, all the textual parts of the document can be treated uniformly. Second, TreeTEX can be trivially ported to any site running TEX. Finally, modularity in the

*This work was supported by a Natural Sciences and Engineering Research Council of Canada Grant A-5692 and a Deutsche Forschungsgemeinschaft Grant Sto167/1-1. It was started during the first author's stay with the Data Structuring Group in Waterloo.

†Institut für Informatik, Universität Freiburg, Rheinstr. 10–12, 7800 Freiburg, West Germany

‡Data Structuring Group, Department of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

description of a tree and TeX's macro capabilities allow for libraries of subtrees and tree classes.

In addition, we have implemented an option that produces drawings which make the *structure* of the trees more obvious to the human eye, even though they may not be as aesthetically pleasing.

1 Aesthetical criteria for drawing trees

One of the most commonly used data structures in computer science is the tree. As many people are using trees in their research or just as illustration tools, they are usually struggling with the problem of *drawing* trees. We are concerned primarily with ordered trees in the sense of [**?**], especially binary and unary-binary trees. A binary tree is a finite set of nodes which either is empty, or consists of a root and two disjoint binary trees called the left and right subtrees of the root. A unary-binary tree is a finite set of nodes which either is empty, or consists of a root and two disjoint unary-binary trees, or consists of a root and one nonempty unary-binary tree. An extended binary tree is a binary tree in which each node has either two nonempty subtrees or two empty subtrees.

For these trees there are some basic agreements on how they should be drawn, reflecting the top-down and left-right ordering of nodes in a tree; see [**?**] and [**?**].

1. Trees impose a distance on the nodes; no node should be closer to the root than any of its ancestors.
2. Nodes of a tree at the same height should lie on a straight line, and the straight lines defining the levels should be parallel.
3. The relative order of nodes on any level should be the same as in the level order traversal of the tree.

These axioms guarantee that trees are drawn as planar graphs: edges do not intersect except at nodes. Two further axioms improve the aesthetical appearance of trees:

4. In a unary-binary tree, each left child should be positioned to the left of its parent, each right child to the right of its parent, and each unary child should be positioned below its parent.

5. A parent should be centered over its children.

An additional axiom deals with the problem of tree drawings becoming too wide and therefore exceeding the physical limit of the output medium:

6. Tree drawings should occupy as little width as possible without violating the other axioms.

In [**?**], Wetherell and Shannon introduce two algorithms for tree drawings, the first of which fulfills axioms 1–5, and the second 1–6. However, as Reingold and Tilford in [**?**] point out, there is a lack of symmetry in the algorithms of Wetherell and Shannon which may lead to unpleasant results. Therefore, Reingold and Tilford introduce a new structured axiom:

7. A subtree of a given tree should be drawn the same way regardless of where it occurs in the given tree.

Axiom 7 allows the same tree to be drawn differently when it occurs as a subtree in different trees. Reingold and Tilford give an algorithm which fulfills axioms 1–5 and 7. Although this algorithm doesn't fulfill axiom 6, the aesthetical improvements are well worth the additional space. Figure **??** illustrates the benefits of axiom 7, and Figure **??** shows that the algorithm of Reingold and Tilford violates axiom 6.

2 The algorithm of Reingold and Tilford

The algorithm of Reingold and Tilford (hereafter called "the RT algorithm") takes a modular approach to the positioning of nodes: The relative positions of the nodes in a subtree are calculated independently from the rest of the tree. After the relative positions of two subtrees have been calculated, they can be joined as siblings in a larger tree by placing them as close together as possible and centering the parent node above them. Incidentally, the modularity principle is the reason that the algorithm fails to fulfill axiom 6; see [**?**]. Two sibling subtrees are placed as close together as possible, during a postorder traversal, as follows. At each node T, imagine that its two subtrees have been drawn and cut out of paper along their contours. Then, starting with the two subtrees superimposed at their roots, move them apart

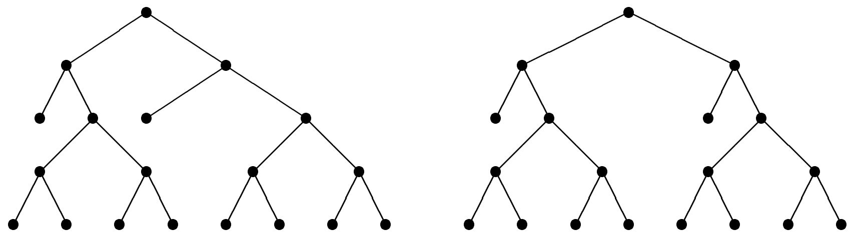


Figure 1: The left tree is drawn by the algorithm of Wetherell and Shannon, and the tidier right one is drawn by the algorithm of Reingold and Tilford.

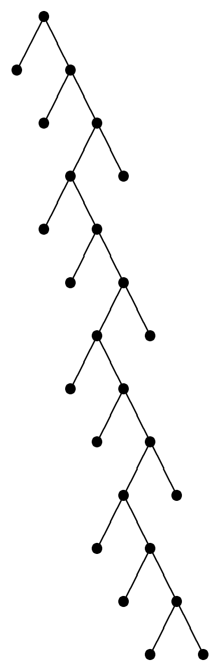


Figure 2: The left tree is drawn by the algorithm of Reingold and Tildford, but the right tree shows that narrower drawings fulfilling all aesthetic axioms are possible.

until a minimal agreed upon distance between the trees is obtained at each level. This can be done gradually: Initially, their roots are separated by some agreed upon minimum distance. Then, at the next lower level, they are pushed apart until the minimum separation is established there. This process is continued at successively lower levels until the bottom of the shorter subtree is reached. At some levels no movement may be necessary; but at no level are the two subtrees moved closer together. When the process is complete, the position of the subtrees is fixed relative to their parent, which is centered over them. Assured that the subtrees will never be placed closer together, the postorder traversal is continued.

A nontrivial implementation of this algorithm has been obtained by Reingold and Tilford that runs in time $O(N)$, where N is the number of nodes of the tree to be drawn. Their crucial idea is to keep track of the contour of the subtrees by special pointers, called threads, such that whenever two subtrees are joined, only the top part of the trees down to the lowest level of the smaller tree need to be taken into account.

The RT algorithm is given in [**?**]. The nodes are positioned on a fixed grid and are considered to have zero width. No labelling is provided. The algorithm only draws binary trees, but is easily extendable to multiway trees.

3 Improving human perception of trees

It is common understanding in book design that aesthetics and readability don't necessarily coincide, and—as Lamport ([**?**]) puts it—books are meant to be read, not to be hung on walls. Therefore, readability is more important than aesthetics.

When it comes to tree drawings, readability means that the structure of a tree must be easily recognizable. This criterion is not always met by the RT algorithm. As an example, there are trees whose structure is very different, the only common thing being the fact that they have the same number of nodes at each level. The RT algorithm might assign identical positions to these nodes making it very hard to perceive the different structures. Hence, we have modified the RT algorithm such that additional white space is inserted between subtrees of *significant* nodes. Here a binary node is called significant if the minimum distance between its two subtrees is taken *below* their root level. Setting the amount of additional white space to zero retains

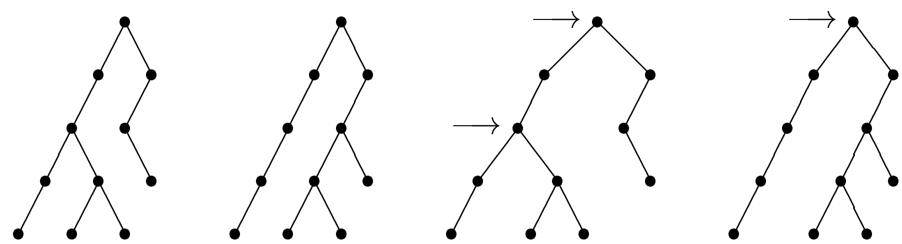


Figure 3: The first two trees get the same placement of their nodes by the RT algorithm, although the structure of the two trees is very different. The alternative drawings highlight the structure of the trees by adding additional white space between the subtrees of ($\longrightarrow$) significant nodes.

the original RT placement. The effect of having nonzero additional white space between the subtrees of significant nodes is illustrated in Figure **??** .

Another feature we have added to the RT algorithms is the possibility to draw an unextended binary tree with the same placement of nodes as its associated extended version. We define the *associated extended version* of a binary tree to be the binary tree obtained by replacing each empty subtree having a nonempty sibling with a subtree consisting of one node. This feature also makes the structure of a tree more prominent; see Figure **??**.

4 Trees in a document preparation environment

Drawings of trees usually don't come alone, but are included in some text which is itself typeset by a text processing system. Therefore, a typical scenario is a pipe of three stages. First comes the tree drawing program which calculates the positioning of the nodes of the tree to be drawn and outputs a description of the tree drawing in some graphics language; next comes a graphics system which transforms this description into an intermediate language which can be interpreted by the output device; and finally comes

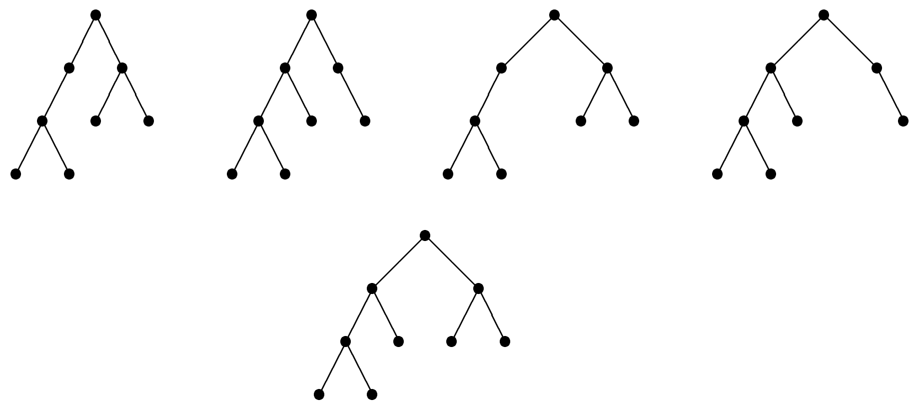


Figure 4: In the first two drawings, the RT algorithm assigns the same placement to the nodes of two trees although their structure is very different. The modified RT algorithms highlights the structure of the trees by optionally drawing them like their extended counterpart, which is given in the second row.

the text processing system which integrates the output of the graphics system into the text.

This scenario loses its linear structure once nodes have to be labelled, since the labelling influences the positioning of the nodes. Labels usually occur inside, to the left of, to the right of, or beneath nodes (the latter only for external nodes), and their extensions certainly should be taken into account by the tree drawing algorithm. But the labels have to be typeset first in order to determine their extensions, preferably by the typesetting program that is used for the regular text, because this method makes for the uniformity in the textual parts of the document and provides the author with the full power of the text processing system for composing the labels. Hence, a more complex communication scheme than a simple pipe is required.

Although a system of two processes running simultaneously might be the most elegant solution, we wanted a system that is easily portable to a large range of hardware at our sites including personal computers with single process operating systems. Therefore, we thought of using a text processing system having programming facilities powerful enough to program a tree drawing algorithm and graphics facilities powerful enough to draw a tree. One text processing system rendering outstanding typographic quality and good enough programming facilities is TeX, developed by Knuth at Stanford University; see [**?**]. The TeX system includes the following programming facilities:

1. datatypes:
   integers (256), dimensions[1] (512), boxes (256), tokenlists (256), boolean variables (unrestricted)

2. elementary statements:
   $a :=$ const, $a := b$ (all types);
   $a := a + b$, $a := a * b$, $a := a/b$ (integers and dimensions);
   horizontal and vertical nesting of boxes

3. control constructs:
   if-then-else statements testing relations between integers, dimensions, boxes, or boolean variables

[1]The term *dimension* is used in TeX to describe physical measurements of typographical objects, like the length of a word.

4. modularization constructs:
   macros with up to 9 parameters (can be viewed as procedures without the concept of local variables).

Although the programming facilities of TEX hardly exceed the abilities of a Turing machine, they are sufficient to handle relatively small programs. How about the graphics facilities? Although TEX has no built-in graphics facilities, it allows the placement of characters in arbitrary positions on the page. Therefore, complex pictures can be synthesized from elementary picture elements treated as characters. Lamport has included such a picture drawing environment in his macro package LATEX, using quarter circles of different sizes and line segments (with and without arrow heads) of different slopes as basic elements; see [**?**]. These elements are sufficient for drawing trees.

This survey of TEX's capabilities implies that TEX may be a suitable text processing system to implement a tree drawing algorithm directly. We are basing our algorithm on the RT algorithm, because this algorithm gives the aesthetically most pleasing results. In the first version presented here, we restrict ourselves to unary-binary trees, although our method is applicable to arbitrary multiway trees. But in order to take advantage of the text processing environment, we expand the algorithm to allow labelled nodes.

In contrast to previous tree drawing programs, we feel no necessity to position the nodes of a tree on a fixed grid. While this may be reasonable for a plotter with a coarse resolution, it is certainly not necessary for TEX, a system that is capable of handling arbitrary dimensions and produces device *independent* output.

5 A representation method for TEXtrees

The first problem to be solved in implementing our tree drawing algorithm is how to choose a good internal representation for trees. A straightforward adaptation of the implementation by Reingold and Tilford requires, for each node, at least the following fields:

1. two pointers to the children of the node

2. two dimensions for the offset to the left and the right child (these may be different once there are labels of different widths to the left and right of the nodes)

3. two dimensions for the x- and y-coordinates of the final position of the nodes

4. three or four labels

5. one token to store the geometric shape (circle, square, framed text etc.) of the node.

Because these data are used very frequently in calculations, they should be stored in registers (that's what variables are called in TEX), rather than being recomputed, in order to obtain reasonably fast performance. This gives a total of $10N$ registers for a tree with N nodes, which would exceed TEX's limited supply of registers. Therefore, we present a modified algorithm hand-tailored to the abilities of TEX. We start with the following observation. Suppose a unary-binary tree is constructed bottom-up, in a postorder traversal. This is done by iterating the following three steps in an order determined by the tree to be constructed.

1. Create a new subtree consisting of one external node.

2. Create a new subtree by appending the two subtrees created last to a new binary node; see Figure **??**.

3. Create a new subtree by appending the subtree created last as a left, right, or unary subtree of a new node; see Figure **??**.

(A pointer to) each subtree that has been created in steps 1–3 is pushed onto a stack, and steps 2 and 3 remove two trees or one, respectively, from the stack before the push operation is carried out. Finally, the tree to be constructed will be the remaining tree on the stack.

This tree traversal is performed twice in the RT algorithm. During the first pass, at each execution of step 2 or step 3, the relative positions of the subtree(s) and of the new node are computed. A closer examination of the RT algorithm reveals that information about the subtree's coordinates is not needed during this pass; the contour information alone would be sufficient.

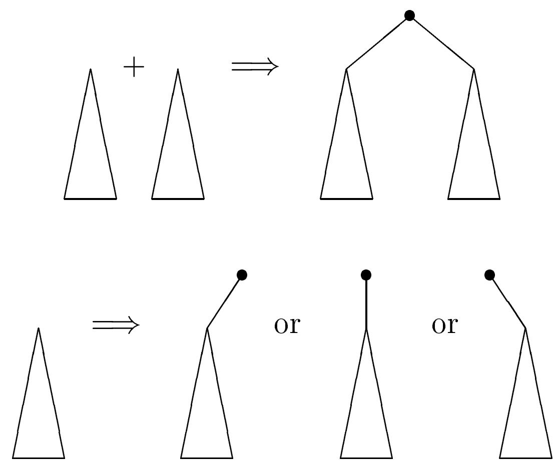


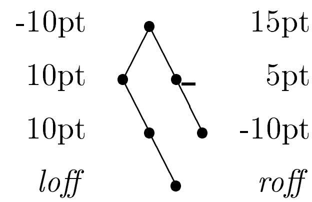
Figure 5: Construction steps 2 and 3

Complete information is only needed in the second traversal, when the tree is actually drawn. Here a special feature of TEX comes in that allows us to save registers. Unlike Pascal, TEX provides the capability of storing a drawing in a single box register that can be positioned freely in later drawings. This means that in our implementation the two passes of the original RT algorithm can be intertwined into a single pass, storing for each subtree on the stack its contour and its drawing. Although the latter is a complex object, it takes only one of TEX's precious registers.

6 The internal representation

Given a tree, the corresponding TEXtree is a box containing the "drawing" of the tree, together with some additional information about the contour of the tree. The reference point of a TEXtree-box is always in the root of the tree. The height, depth, and width of the box of a TEXtree are of no importance in this context.

The additional information about the contour of the tree is stored in some registers for numbers and dimensions and is needed in order to put subtrees together to form a larger tree. *loff* is an array of dimensions which contains

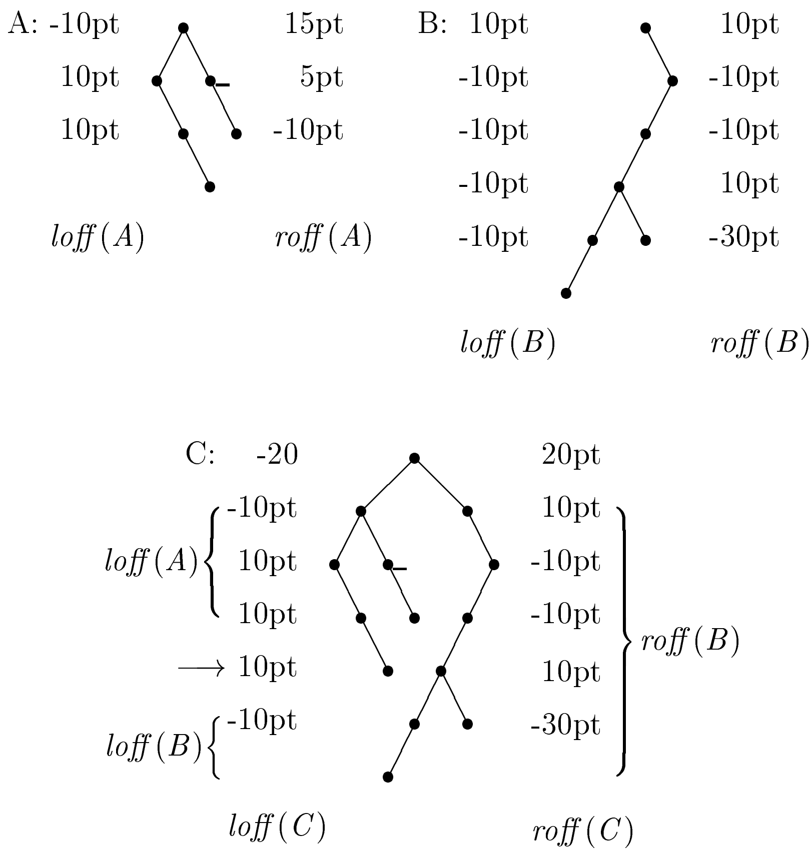


height: 3, type: dot, ltop: 2pt, rtop: 2pt, lmoff: -10pt, rmoff: 20pt, lboff: 10pt, rboff: 10pt.

Figure 6: A TEXtree consists of the drawing of the tree and the additional information. The width of the dots is 4pt, the minimal separation between adjacent nodes is 16pt, making for a distance of 20pt center to center. The length of the small rule labelling one of the nodes is 5pt. The column left (right) of the tree drawing is the array *loff* (*roff*), describing the left (right) contour of the tree. At each level, the dimension given is the horizontal offset between the border at the current and at the next level. The offset between the left border of the root node and the leftmost node at level 1 is -10pt, the offset between the right border of the root node and the rightmost node at level 1 is 15pt, etc.

for each level of the tree the horizontal offset between the left end of the leftmost node at the current level and the left end of the leftmost node at the next level. *lmoff* holds the horizontal offset between the root and the leftmost node of the whole tree. *lboff* holds the horizontal offset between the root and the leftmost node at the bottom level of the tree. Finally, *ltop* holds the distance between the reference point of the tree and the leftmost end of the root. The same is true for *roff*, *rmoff*, *rboff*, and *rtop*; just replace "left" by "right". Finally, *height* holds the height of the tree, and *type* holds the geometric shape of the root of the tree. Figure **??** shows an example TEXtree, i.e. a tree drawing and the corresponding additional information.

Given two TEXtrees A and B, how can a new TEXtree C be built that consists of a new root and has A and B as subtrees? An example is given in Figure **??**.



	A	B	C
height	3	5	6
type	dot	dot	dot
ltop	2pt	2pt	2pt
rtop	2pt	2pt	2pt
lmoff	-10pt	-30pt	-30pt
rmoff	20pt	10pt	30pt
lboff	10pt	-30pt	-10pt
rboff	10pt	-30pt	-10pt

level	*totsep*	*currsep*
0	20pt	0/16pt
1	25pt	11/16
2	40pt	1/16pt
3	40pt	16pt

Figure 7: The TEXtrees A and B are combined to form the larger TEXtree C. The small table gives the history of computation for *totsep* and *currsep*.

First we determine which tree is higher; this is B in the example. Then we have to compute the minimal distance between the roots of A and B, such that at all levels of the trees there is free space of at least *minsep* between the trees when they are drawn side by side. For this purpose we keep track of two values, *totsep* and *currsep*. The variables *totsep* and *currsep* hold the total distance between the roots and the distance between the rightmost node of A and the leftmost node of B at the current level. In order to calculate *totsep* and *currsep*, we start at level 0 and visit each level of the trees until we reach the bottom level of the smaller tree; this is A in our example.

At level 0, the distance between the roots of A and B should be at least *minsep*. Therefore, we set $totsep := minsep + rtop(A) + ltop(B)$ and $currsep := minsep$. Using $roff(A)$ and $loff(B)$, we can proceed to calculate *currsep* for the next level. If $currsep < minsep$, we have to increase *totsep* by the difference and update *currsep*. This process is iterated until we reach the lowest level of A. Then *totsep* holds the final distance between the nodes of A and B, as calculated by the RT algorithm. If the root of C is a significant node, then the additional space , which is 0pt by default, is added to *totsep*. However, the approach of synthesizing drawings from simple graphics characters allows only a finite number of orientations for the tree edges; therefore, *totsep* must be increased slightly to fit the next orientation available.

Now we are ready to construct the box of TEXtree C. Simply put A and B side by side, with the reference points *totsep* units apart, insert a new node above them, and connect the parent and children by edges.

Next, we update the additional information for C. This can be done by using the additional information for A and B. Note that most components of $roff(C)$ and $lroff(C)$ are the same as in the higher tree, which is B in our case. So, if we can avoid moving this information around, we only have to access $height(A) + const$ many counters in order to update the additional information for C. This implies that we can apply the same argument as in [**?**], which gives us a running time of $O(N)$ for drawing a tree with N nodes.

Therefore, we must carefully design the storage allocation for the additional information of TEXtrees in order to fulfill the following requirements: If a new tree is built from two subtrees, the additional information of the new tree should share storage with its larger subtree. Organizational overhead, that is, pointers which keep track of the locations of different parts of additional information, must be avoided. This means that all the additional information for one TEXtree should be stored in a row of consecutive dimension registers such that only one pointer granting access to the first element in this row is needed. On the other hand, each parent tree is higher and therefore needs more storage than its subtrees. So we must ensure that there is always enough space in the row for more information.

The obvious way to fulfill these requirements is to use a stack and to allow only the topmost TEXtrees of this stack to be combined into a larger tree at any time. This leads to the following register allocation: A subsequent number of box registers contains the treeboxes of the subtrees in the stack. A subsequent number of token registers contains the type information for the nodes of the subtrees in the stack. For each subtree in the stack, a subsequent number of dimension registers contains the contour information of the subtree. The ordering of these groups of dimension registers reflects the ordering of the subtrees in the stack. Finally, a subsequent number of counter registers contains the height and the address of the first dimension register for each subtree in the stack. Four address counters store the addresses of the last treebox, type information, height, and address of contour information. A sketch of the register organization for a stack of TEXtrees is provided in Figure **??**.

When a new node is pushed onto the stack, the treebox, type information, height, address of contour information, and contour information are stored in the next free registers of the appropriate type, and the four address counters are updated accordingly.

When a new tree is formed from the topmost subtrees on the stack, the treebox, type information, height, and address of contour information of the new tree are sorted in the registers formerly used by the bottommost subtree that has occured in the construction step, and the four address registers are updated accordingly. This means that these informations for the subtrees are no longer accessible. The contour information of the new subtree is stored in the same registers as the contour information of the larger subtree used in the construction, apart from the left and right offset of the root to the left and right child, which are stored in the following dimension registers. That means that gaps can occur between the contour information of subsequent subtrees in the stack, namely when the right subtree, which is on a higher position on the stack, is higher than the left one. In order to avoid these gaps, the user can specify an option `\lefttop` when entering a binary node, which makes the topmost tree in the stack the left subtree of the node.

Dimension registers
$lmoff(1)$ $rmoff(1)$ $lboff(1)$ $rboff(1)$ $ltop(1)$ $rtop(1)$
$loff(h_1)$ $roff(h_1)$ $\ldots$ $loff(1)$ $roff(1)$
$\ldots$
$lmoff(n)$ $rmoff(n)$ $lboff(n)$ $rboff(n)$ $ltop(n)$ $rtop(n)$
$loff(h_n)$ $roff(h_n)$ $\ldots$ $loff(1)$ $roff(1)$

Counter registers
lasttreebox lasttreeheight lasttreeinfo lasttreetype
$treeheight(1)$ $diminfo(1)$ $\ldots$ $treeheight(n)$ $diminfo(n)$

Box registers
$treebox(1)$ $\ldots$ $treebox(n)$

Token registers
$type(1)$ $\ldots$ $type(n)$

Figure 8: *lasttreebox*, *lasttreeheight*, *lasttreeinfo*, *lasttreetype* contain pointers to $treebox(n)$ $treeheight(n)$, $lmoff(n)$, $type(n)$, $diminfo(i)$ contains a pointer to $lmoff(i)$. Unused dimension registers are allowed between the dimension registers of subsequent trees. The counter registers *lasttreebox*,...,$diminfo(n)$ serve as a directory mechanism to access the TEXtrees on the stack.

This stack concept also has consequences for the design of the user interface that is discussed in Section **??**.

7 Space cost analysis

Suppose we want to draw a unary-binary tree T of height h having N nodes[2]. According to our internal representation, for each subtree in the stack we need

1. one box register to store the box of the TEXtree.
2. one token register to store the type of the root of the subtree.
3. $2h' + 6$ dimension registers to store the additional information, where h' is the height of the subtree.
4. three counter registers to store the register numbers of the box register, the token register, and the first dimension register above.

The following lemma relates to h and N the number of subtrees of T which are on the stack simultaneously and their heights.

Lemma 7.1

1. *At any time, there are at most $h + 1$ subtrees of T on the stack.*
2. *For each set $\mathcal{T}$ of subtrees of T which are on the stack simultaneously we have*
$$\sum_{T' \in \mathcal{T}} (\mathrm{ht}(T') + 1) \leq \min(N, \frac{(h+1)(h+2)}{2}).$$

Proof

1. By induction on h.

[2]The height h and the number of nodes N refer to the drawing of the tree. N is the number of circles, squares etc. actually drawn, and h is the number of levels in the drawing minus 1.

2. The trees in $\mathcal{T}$ are pairwise disjoint, and each tree of height h' has at least $h'+1$ nodes. This implies

$$\sum_{T' \in \mathcal{T}} (\mathrm{ht}(T')+1) \leq N.$$

The second part is shown by induction on h. The basis $h=0$ is clear. Assume the assumption holds for all trees of height less than h. If $\mathcal{T}$ contains only subtrees of either the left or the right subtree of T, we have

$$\sum_{T' \in \mathcal{T}} (\mathrm{ht}(T')+1) \leq \frac{h(h+1)}{2} \leq \frac{(h+1)(h+2)}{2}.$$

Otherwise, $\mathcal{T}$ contains the left or the right subtree T_s of T. Then all elements of $\mathcal{T} - \{T_s\}$ belong to the other subtree. This implies

$$\begin{aligned} \sum_{T' \in \mathcal{T}} (\mathrm{ht}(T')+1) &\leq \mathrm{ht}(T_s)+1+\sum_{T' \in \mathcal{T}-\{T_s\}} (\mathrm{ht}(T')+1) \\ &\leq h+\frac{h(h+1)}{2} \leq \frac{(h+1)(h+2)}{2}. \qquad \square \end{aligned}$$

Therefore, our implementation uses at most $9h + 2\min(N, (h+1)(h+2)/2)$ registers. In order to compare this with the $10N$ registers used in the straightforward implementation, an estimation of the average height of a tree with N nodes is needed. Several results, depending on the type of trees and of the randomization model, are cited in Figure **??**, which compares the number of registers used in a straightforward implementation with the average number of registers used in our implementation. This table shows clearly the advantage of our implementation.

8 The user interface

8.1 General design considerations

The user interface of TreeTEX has been designed in the spirit of the thorough separation of the logical description of document components and their layout; see [**?**, **?**]. This concept ensures both uniformity and flexibility of document layout and frees authors from layout problems which have nothing

nodes	registers (straight-forward)	average registers		
		extended binary trees ($\sqrt{\pi n}$) [?]	unary-binary trees ($\sqrt{3\pi n}$) [?]	binary search trees ($4.311 \log n$) [?]
8	80	61.12	94.15	51.04
9	90	65.86	100.89	55.02
10	100	70.44	107.37	58.80
11	110	74.91	113.64	62.41
12	120	79.26	119.71	65.87
20	200	111.34	163.56	90.48
30	300	147.37	211.33	117.31
40	400	180.89	254.75	132.58
50	500	212.80	295.37	143.54

Figure 9: The numbers of registers used by a straightforward implementation (second column) and by our modified implementation (third to fifth column) of the RT algorithm are given for different types of trees and randomization models. The formula in parentheses indicates the average height of the respective class of trees, as depending on the number of nodes.

to do with the substance of their work. For some powerful implementations and projects see [?, ?, ?, ?, ?].

In this context, the description of a tree is given in a purely logical form, and layout variations are defined by a separate style command which is valid for all trees of a document.

A second design principle is to provide defaults for all specifications, thereby allowing the user to omit many definitions if the defaults match what he or she wants.

The node descriptions of a tree must be entered in postorder. This fits the internal representation of TeXtrees best. Although this is a natural method of describing a tree, a user might prefer more flexible description methods. However, note that instances of well defined tree classes can be described easily by TeX macros. In section ??. we give examples of macros for complete binary trees and Fibonacci trees.

TreeTeX uses the picture making macros of LaTeX. If TreeTeX is used with any other macro package or format, the picture macros of LaTeX are included automatically.

8.2 The description of a tree

The description of a tree is started by the command `\beginTree` and closed by `\endTree` (or `\begin{Tree}` and `\end{Tree}` in LaTeX). The description can be started in any mode; it defines a box and two dimensions. The box is stored in the box register `\TeXTree` and contains the drawing of the tree. The box has zero height and width, and its depth is the height of the drawing. The reference point is in the center of the node of the tree. The dimensions are stored in the registers `\leftdist` and `\rightdist` and describe the distance between the reference point and the left and right margin of the drawing. These data can be used to position the drawing of the tree.

Note that the TreeTeX macros don't contribute anything to the current page but only store their results in the registers `\TeXTree`, `\leftdist`, and `\rightdist`. It is the user's job to put the drawing onto the page, using the commands `\copy` or `\box` (or `\usebox` in LaTeX).

Each matching pair of `\beginTree` and `\endTree` must contain the description for only *one* tree. Descriptions of trees cannot be nested and new registers cannot be allocated inside a matching pair of `\beginTree` and `\endTree`.

As already stated, each tree description defines the nodes of the tree in postorder, that is, a tree description is a particular sequence of node descriptions.

A node description, in turn, consists of the macro \node, followed by a list of node options, included in braces. The list of node options may be empty. The node options describe the labels, the geometric shape (type), and the outdegree of the node. Default values are provided for all options which are not explicitly specified. The following node options are available:

1. \lft{<label>}, \rght{<label>}, \cntr{<label>}, \bnth{<label>}:
   These options describe the labels which are put to the left of, to the right of, in the center of, or beneath the node (the latter only makes sense for external nodes). The arguments of these macros are processed in internal horizontal mode (LR-mode in LaTeX), but can consist of arbitrary nested boxes for more sophisticated labels. For each of these options, the default is an empty label.

2. \external, \unary, \leftonly, \rightonly:
   These options describe the outdegree of the node. The default is binary (no outdegree option is specified).

3. \type{<type>}:
   This option describes the type or geometric shape of the node. <type> can have the values square, dot, text, or frame. The default value is circle (no type is specified). A node of type square has a fixed width, while a node of type frame has its width determined by the center label. A node of type text has no frame around its center label. The center label can have arbitrary width.

3. \leftthick, \rightthick: These options change the thickness of the left or right outgoing edge of a binary node. Defaults are thin edges (neither option is specified).

4. \lefttop:
   The node option \lefttop in a binary node makes the last entered subtree the left child of the node (the right child is the default). This option helps to cut down on the number of dimension registers used

during the construction of a tree. As a rule of thumb, this option is recommended when the left subtree has a smaller height than the right subtree, that is, in this case the right subtree should be entered before the left one and their parent should be assigned the option `\lefttop`.

8.3 Macros for classes of trees

Tree descriptions can be produced by macros. This is especially useful for trees which belong to a larger class of trees and which can be specified by some simple parameters. A small library of such macros is provided in the file `TreeClasses.tex`.

1. `\treesymbol{<node options>}`:
   This macro produces a triangular tree symbol which can be included in a tree description instead of an external node. Labels for these tree symbols are described as for ordinary nodes. In addition, the options `\lvls{<number>}` and `\slnt{<number>}` are provided. `\lvls` defines the number of levels in the tree over which the triangle extends, and `\slnt` gives the slant of the sides of the triangle, ranging from 1 (minimal) to 24 (maximal). On the other hand, `\treesymbol` does not expand to a tree description, because a tree symbol cannot be built from subtrees, and, on the other hand, it is not a node, because it is allowed to extend over several tree levels and therefore has a longer contour than an ordinary node.

2. `\binary{<bin specification>}`:
   This macro truly expands to a tree description. It produces a complete binary tree, that is, an extended binary tree, where, for a given h, all external nodes appear at level h or $h-1$, and all external nodes at level h lie left of those at level $h-1$. `<bin specification>` consists of the following options: `\no{<number>}` defines the number of internal nodes, with `<number>` greater than 0, and `\squareleaves` produces leaves of type `square`. Defaults are `\no{1}` and leaves of type `circle`.

3. `fibonacci{<fib specification>}`:
   This macro produces a Fibonacci tree. `<fib specification>` allows for the three options `\hght{<number>}`, `\unarynodes`, and `\squareleaves`.

Normally, a Fibonacci tree of height $h+2$ is a binary tree with Fibonacci trees of height h and $h+1$ as left and right subtrees. The option \unarynodes means that the Fibonacci tree is augmented by unary nodes such that each two subtree siblings have the same height. These are examples of what has been called brother-trees in the literature; see [?]. Defaults are \hght{0}, the unaugmented version of a Fibonacci tree, and external nodes of type circle.

8.4 Style options for trees

The TreeTEX package includes a style command \Treestyle{<style option>}, where <style option> contains all the parameter settings the user might want to change. Normally, the command \Treestyle appears only once at the beginning of the document and the style options are valid for all trees of the document.

The changes in the style options are global. A \Treestyle command changes only the specified style options; non-specified options retain the last specified value or the default value, respectively. The following style options are available:

1. \treefonts{<font options>}:
   \treefonts is invoked by \beginTree, and it simply executes whatever is specified in <font options>. Defaults are \treefonts{\tenrm} (or \treefonts{\normalsize\rm} in LaTEX).

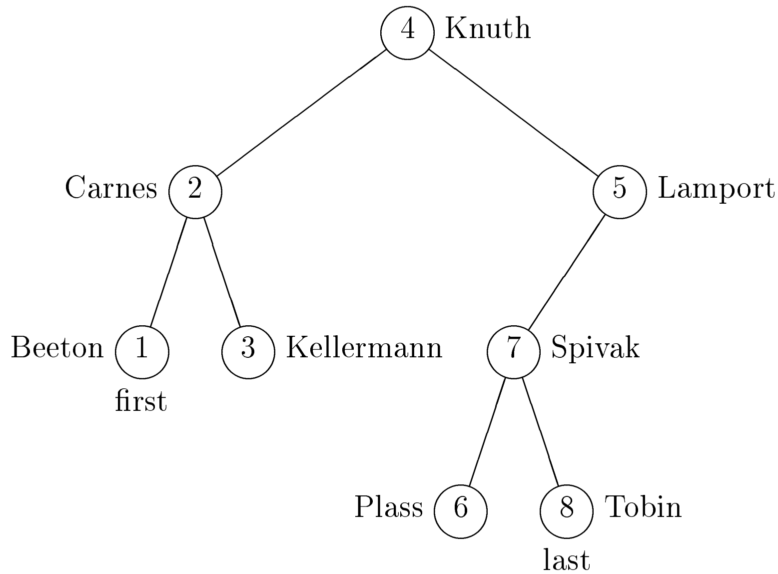
2. \nodesize{<size>}:
   \nodesize defines the size of the nodes. <size> is a dimension and specifies the diameter of circle nodes. The width of square nodes is adjusted accordingly to be slightly smaller than the diameter of circle nodes in order to balance their appearance. Furthermore, \nodesize adjusts the amount of space by which the baseline of the labels is placed beneath the center of the node. The default value of \nodesize suits the default of \treefonts (taking into account the size option of LaTEX's document style).

3. \vdist{<dimen>}, \minsep{<dimen>}, \addsep{<dimen>}:
   vdist specifies the vertical distance between two subsequent levels of

the tree. Default is `\vdist{60pt}`. `\minsep` specifies the minimal horizontal distance between two adjacent nodes. Default is `\minsep{20pt}`. `\addsep` specifies the additional amount of horizontal space by which two subtree siblings are pushed apart farther than calculated by the RT algorithm, if the level at which they are closest is beneath their root level. Default is `\addsep{0pt}`

4. `\extended`, `\nonextended`:
   With the option `\extended` in effect, the nodes of a binary tree are placed in exactly the same way as they would be in the associated extended version of the tree (the missing nodes are assumed to have no labels). The default is `\nonextended`, that is the usual layout.

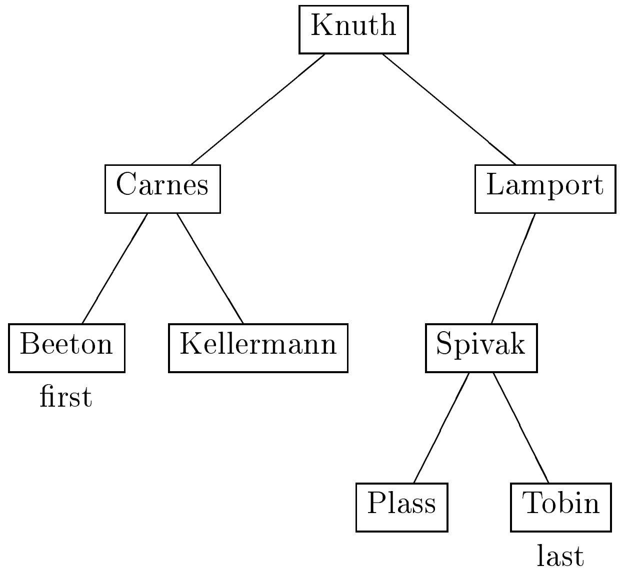
Some examples of tree descriptions are given in the next figures. A detailed description of the TreeTEX macros is given in [**?**].




```
\begin{Tree}
\node{\external\bnth{first}\cntr{1}\lft{Beeton}}
\node{\external\cntr{3}\rght{Kellermann}}
\node{\cntr{2}\lft{Carnes}}
\node{\external\cntr{6}\lft{Plass}}
\node{\external\bnth{last}\cntr{8}\rght{Tobin}}
\node{\cntr{7}\rght{Spivak}}
\node{\leftonly\cntr{5}\rght{Lamport}}
\node{\cntr{4}\rght{Knuth}}
\end{Tree}

\hspace{\leftdist}\usebox{\TeXTree}\hspace{\rightdist}
```

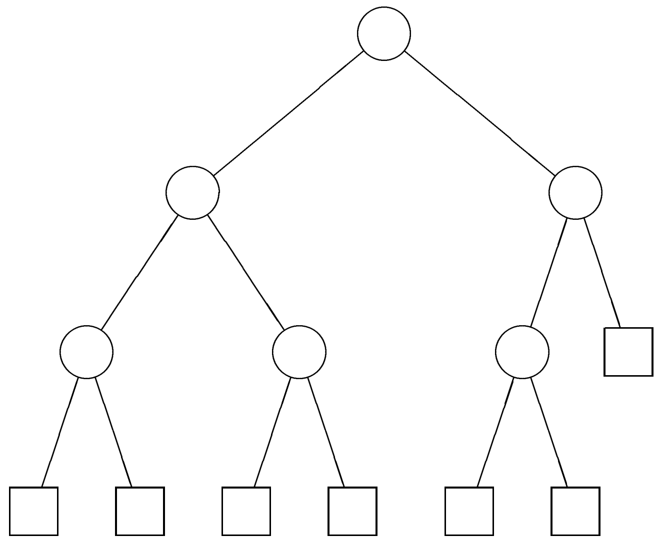
Figure 10: This is an example of a tree that includes labels.




```
\begin{Tree}
\node{\external\type{frame}\bnth{first}\cntr{Beeton}}
\node{\external\type{frame}\cntr{Kellermann}}
\node{\type{frame}\cntr{Carnes}}
\node{\external\type{frame}\cntr{Plass}}
\node{\external\type{frame}\bnth{last}\cntr{Tobin}}
\node{\type{frame}\cntr{Spivak}}
\node{\leftonly\type{frame}\cntr{Lamport}}
\node{\type{frame}\cntr{Knuth}}
\end{Tree}

\hspace{\leftdist}\usebox{\TeXTree}\hspace{\rightdist}
```

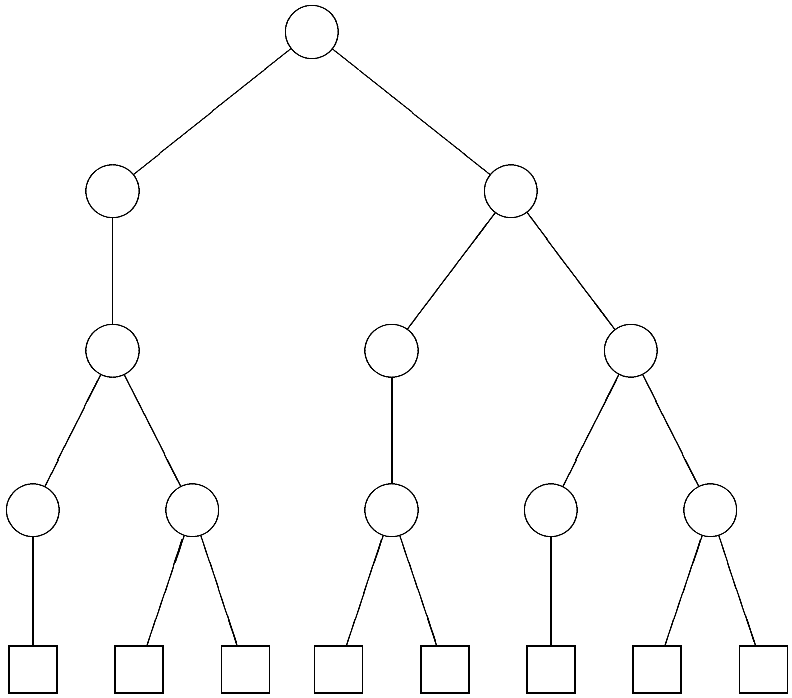
Figure 11: This is an example of a tree with framed center labels.



```
\begin{Tree}
\binary{\no{6}\squareleaves}
\end{Tree}

\hspace{\leftdist}\usebox{\TeXTree}\hspace{\rightdist}
```

Figure 12: This is an example of a complete binary tree.



```
\begin{Tree}
\fibonacci{\hght{4}\unarynodes\squareleaves}
\end{Tree}
\hspace{\leftdist}\usebox{\TeXTree}\hspace{\rightdist}
```

Figure 13: This is an example of a Fibonacci tree.

References

[1] R. J. Beach. *Setting Tables and Illustrations with Style.* PhD thesis, University of Waterloo, 1985.

[2] A. Brüggemann-Klein, P. Dolland, and A. Heinz. How to please authors and publishers: a versatile document preparation system at Karlsruhe. In J. Désarménien, editor, *TEX for Scientific Documentation*, Strasbourg, France, June 1986. LNCS 236.

[3] A. Brüggemann-Klein and D. Wood. *TreeTEX: Documentation and User Handbook.* Technical Report, University of Waterloo, 1987.

[4] N.G. de Bruijn, D. Knuth, and S.O. Rice. The average height of planted plane trees. In R.C. Read, editor, *Graph Theory and Computing*, 1972.

[5] L. Devroye. A note on the height of binary search trees. *Journal of the ACM*, 33(3), July 1986.

[6] Ph. Flajolet and A. Odlyzko. The average height of binary trees and other simple trees. *Journal of Computer and System Sciences*, 25, 1982.

[7] R. Furuta, J. Scofield, and A. Shaw. Document formatting systems: surveys, concepts, issues. *Computing Surveys*, 14(3), 1982.

[8] Ch. F. Goldfarb. A generalized approach to document markup. *SIGPLAN Notices of the ACM*, June 1981.

[9] D. E. Knuth. *Fundamental Algorithms.* Volume 1 of *The Art of Computer Programming*, Addison-Wesley, Reading, Massachusetts, 1973.

[10] D. E. Knuth. *The TEXbook.* Volume A of *Computers & Typesetting*, Addison-Wesley, Reading, Massachusetts, 1986.

[11] L. Lamport. *LATEX, User's Guide & Reference Manual.* Addison-Wesley, Reading, Massachusetts, 1986.

[12] Th. Ottmann, H.-W. Six. Eine neue Klasse ausgeglichener Binärbäume. *Angewandte Informatik*, 9, 1976.

[13] V. Quint, I. Vatton, and H. Bedor. Grif: an interactive environment for TEX. In J. Désarménien, editor, *TEX for Scientific Documentation*, Strasbourg, France, June 1986. LNCS 236.

[14] B. K. Reid. *Scribe: A Document Specification Language and its Compiler*. PhD thesis, Carnegie Mellon University, 1980.

[15] E. M. Reingold and J. S. Tilford. Tidier drawings of tree. *IEEE Transactions on Software Engineering*, 7(2), March 1981.

[16] K. J. Supowit and E. M. Reingold. The complexity of drawing trees nicely. *Acta Informatica*, 18, 1983.

[17] Ch. Wetherell and A. Shannon. Tidy drawings of trees. *IEEE Transactions on Software Engineering*, 5(5), September 1979.