Introduction to Algorithms

Binary Trees

Outline

In this talk, we will look at the binary tree data structure:

- Definition
- Properties
- A few applications
  - Ropes (strings)
  - Expression trees

Definition

This is not a binary tree:

Definition

The arbitrary number of children in general trees is often unnecessary—many real-life trees are restricted to two branches

- Expression trees using binary operators
- An ancestral tree of an individual, parents, grandparents, *etc*.
- Phylogenetic trees
- Lossless encoding algorithms

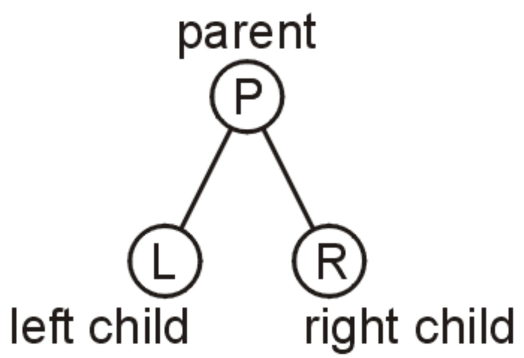
There are also issues with general trees:

- There is no natural order between a node and its children

Definition

A binary tree is a restriction where each node has exactly two children:

- Each child is either empty or another binary tree
- This restriction allows us to label the children as *left* and *right* subtrees

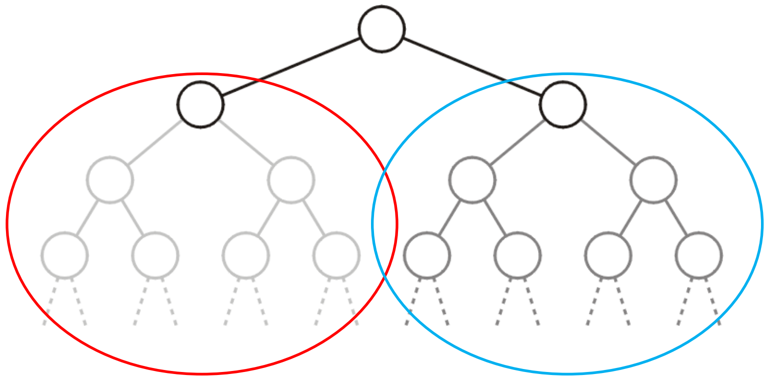


At this point, recall that $\lg(n) = \Theta(\log_b(n))$ for any b

Definition

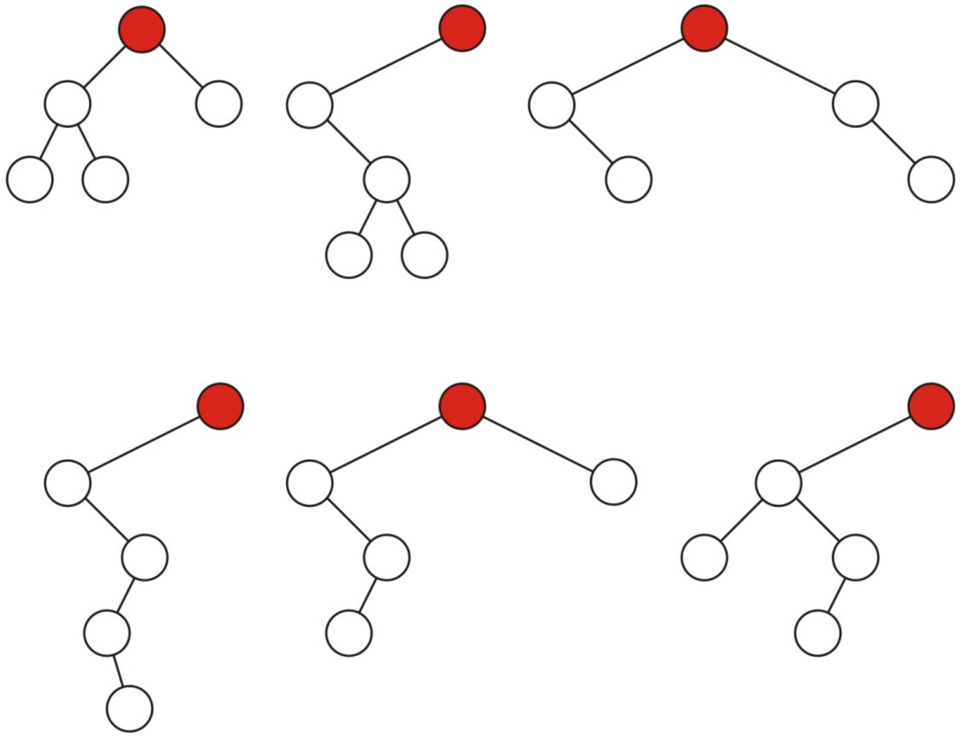
We will also refer to the two sub-trees as

- The left-hand sub-tree, and
- The right-hand sub-tree



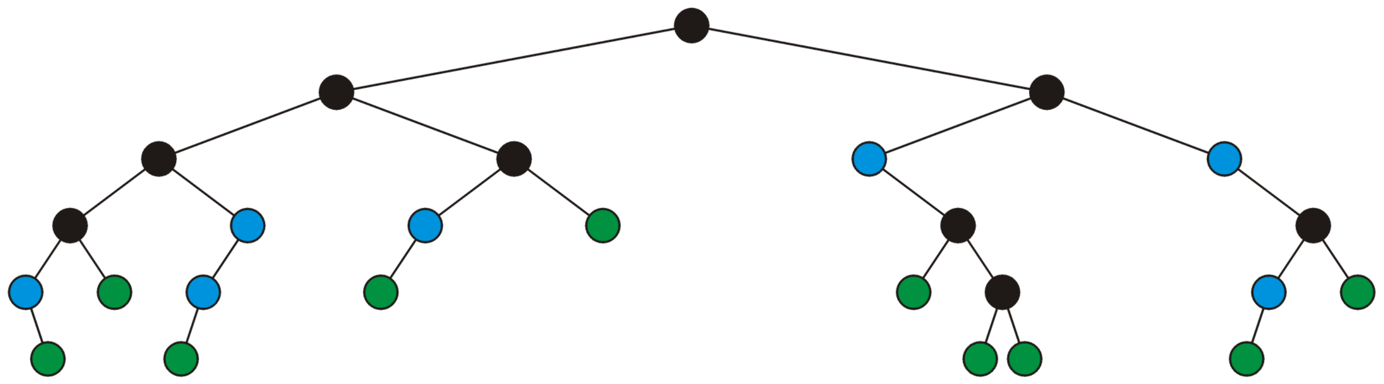
Definition

Sample variations on binary trees with five nodes:



Definition

A *full* node is a node where both the left and right sub-trees are non-empty trees

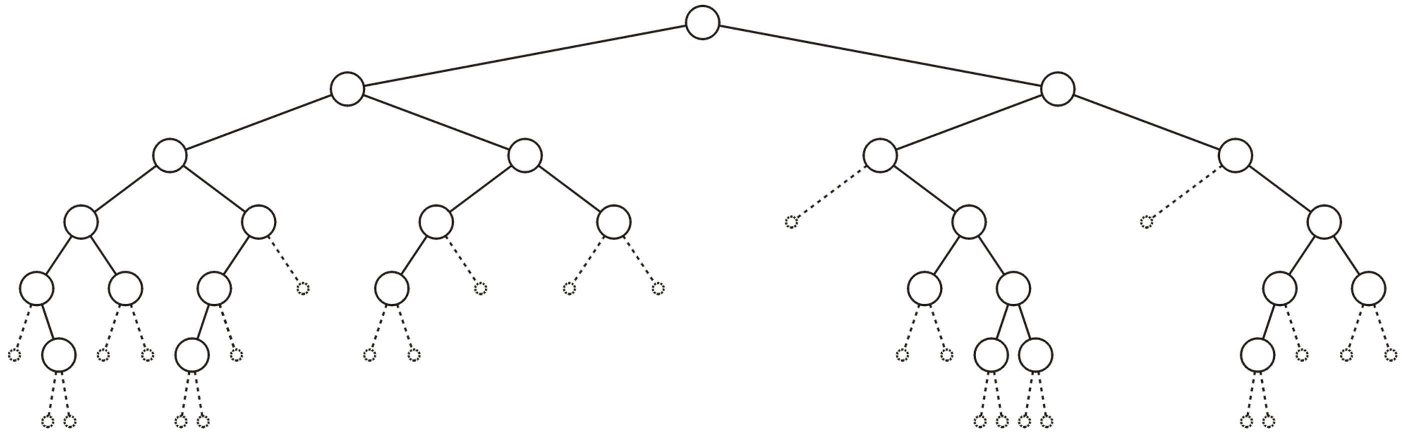


Legend:

●	●	●
full nodes	neither	leaf nodes

Definition

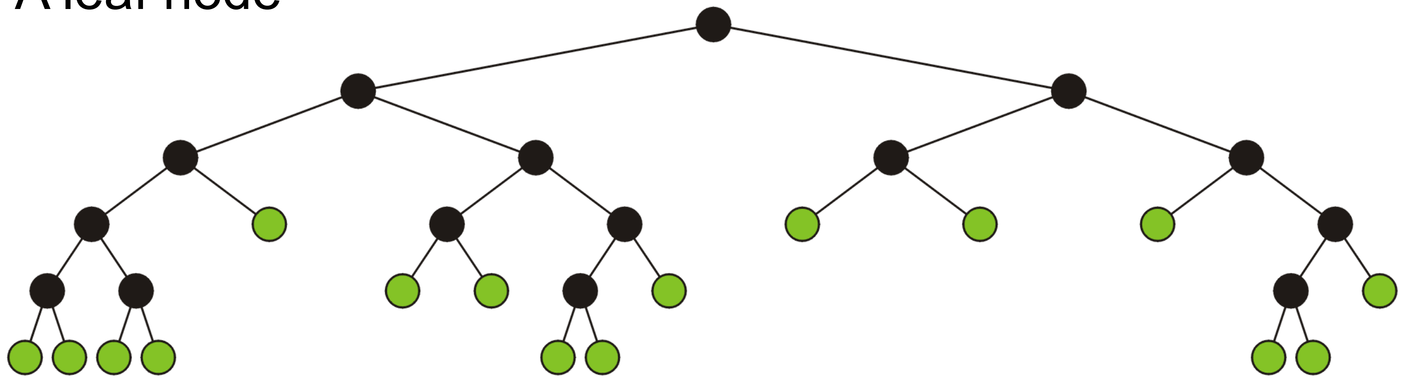
An *empty node* or a *null sub-tree* is any location where a new leaf node could be appended



Definition

A *full binary tree* is where each node is:

- A full node, or
- A leaf node



These have applications in

- Expression trees
- Huffman encoding

Binary Node Class

The binary node class is similar to the single node class:

```
#include <algorithm>

template <typename Type>
class Binary_node {
    protected:
        Type node_value;
        Binary_node *p_left_tree;
        Binary_node *p_right_tree;

    public:
        Binary_node( Type const & );

        Type value() const;
        Binary_node *left() const;
        Binary_node *right() const;

        bool is_leaf() const;
        int size() const;
        int height() const;
        void clear();
}
```

Binary Node Class

We will usually only construct new leaf nodes

```
template <typename Type>
Binary_node<Type>::Binary_node( Type const &obj ):
node_value( obj ),
p_left_tree( nullptr ),
p_right_tree( nullptr ) {
    // Empty constructor
}
```

Binary Node Class

The accessors are similar to that of `Single_list`

```
template <typename Type>
Type Binary_node<Type>::value() const {
    return node_value;
}

template <typename Type>
Binary_node<Type> *Binary_node<Type>::left() const {
    return p_left_tree;
}

template <typename Type>
Binary_node<Type> *Binary_node<Type>::right() const {
    return p_right_tree;
}
```

Binary Node Class

Much of the basic functionality is very similar to `Simple_tree`

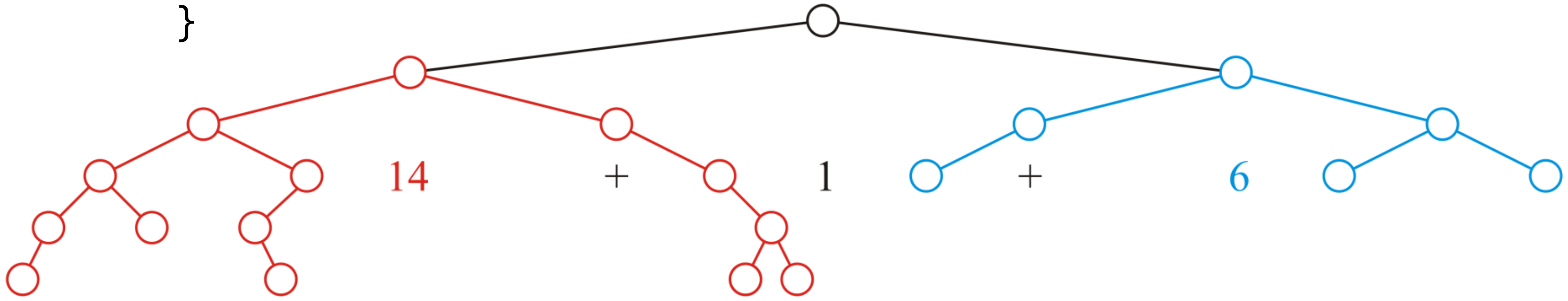
```
template <typename Type>
bool Binary_node<Type>::is_leaf() const {
    return (left() == nullptr) && (right() == nullptr);
}
```

Size

The recursive size function runs in $\Theta(n)$ time and $\Theta(h)$ memory

- These can be implemented to run in $\Theta(1)$

```
template <typename Type>
int Binary_node<Type>::size() const {
    if ( left() == nullptr ) {
        return ( right() == nullptr ) ? 1 : 1 + right()->size();
    } else {
        return ( right() == nullptr ) ?
            1 + left()->size() :
            1 + left()->size() + right()->size();
}
```

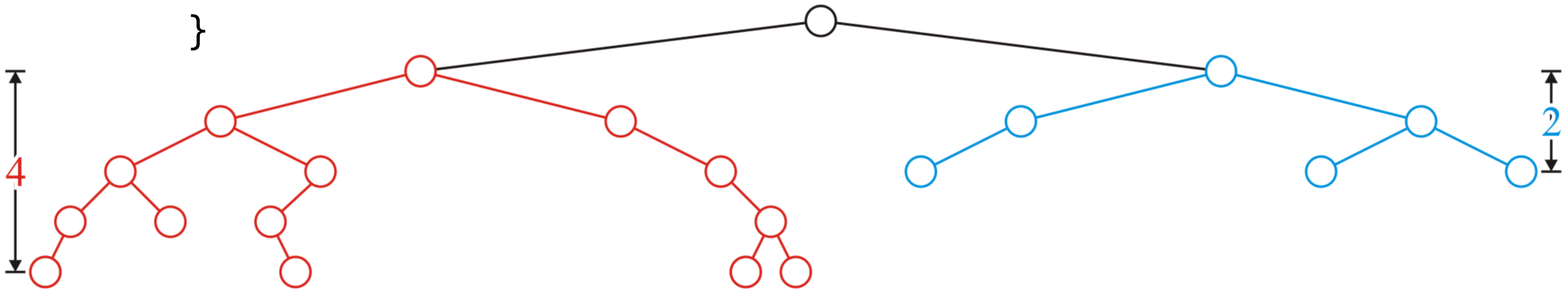


Height

The recursive height function also runs in $\Theta(n)$ time and $\Theta(h)$ memory

- Later we will implement this in $\Theta(1)$ time

```
int Binary_node<Type>::height() const {
    if ( left() == nullptr ) {
        return ( right() == nullptr ) ? 0 : 1 + right()->height();
    } else {
        return ( right() == nullptr ) ?
            1 + left()->height() :
            1 + left()->height() + right()->height();
}
```



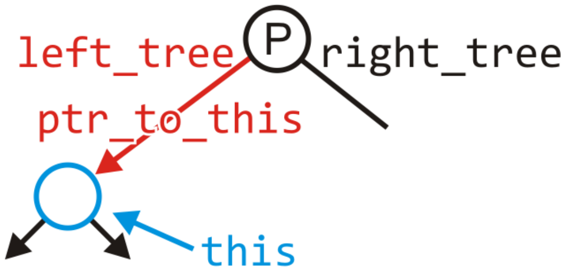
Clear

Removing all the nodes in a tree is similarly recursive:

```
template <typename Type>
void Binary_node<Type>::clear( Binary_node *&p_to_this ) {
    if ( left() != nullptr ) {
        left()->clear( p_left_tree );
    }

    if ( right() != nullptr ) {
        right()->clear( p_right_tree );
    }

    delete this;
    p_to_this = nullptr;
}
```



Run Times

Recall that with linked lists and arrays, some operations would run in $\Theta(n)$ time

The run times of operations on binary trees, we will see, depends on the height of the tree

We will see that:

- The worst is clearly $\Theta(n)$
- Under average conditions, the height is $\Theta\left(\sqrt{n}\right)$
- The best case is $\Theta(\ln(n))$

Run Times

If we can achieve and maintain a height $\Theta(\lg(n))$, we will see that many operations can run in $\Theta(\lg(n))$ we

Logarithmic time is not significantly worse than constant time:

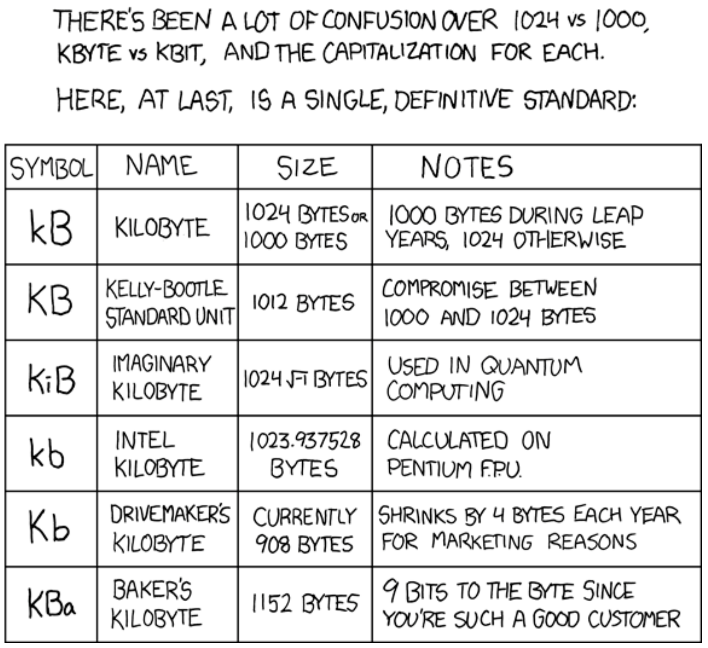
$$\lg(\ 1000\ ) \approx 10 \qquad \text{kB}$$

$$\lg(\ 1\ 000\ 000\ ) \approx 20$$

$$\lg(\ 1\ 000\ 000\ 000\ ) \approx 30$$

$$\lg(\ 1\ 000\ 000\ 000\ 000\ ) \approx 40$$

$$\lg(\ 1000^n\ ) \approx 10\ n$$



THERE'S BEEN A LOT OF CONFUSION OVER 1024 vs 1000, KBYTE vs KBIT, AND THE CAPITALIZATION FOR EACH.

HERE, AT LAST, IS A SINGLE, DEFINITIVE STANDARD:

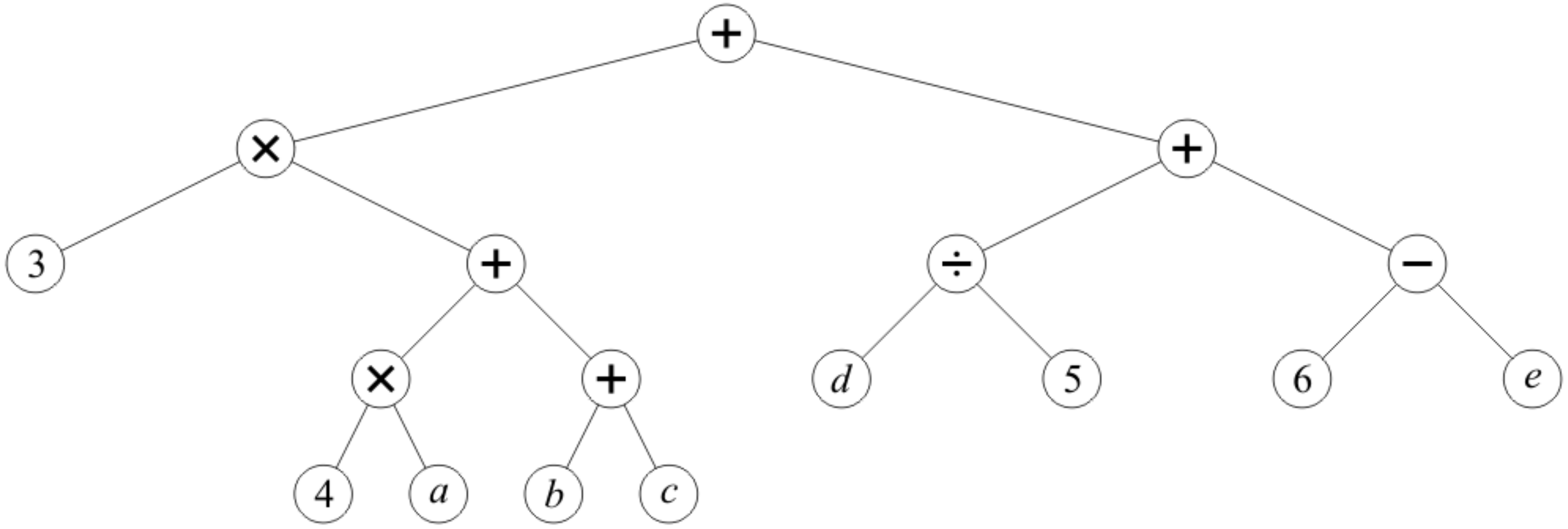
SYMBOL	NAME	SIZE	NOTES
kB	KILOBYTE	1024 BYTES OR 1000 BYTES	1000 BYTES DURING LEAP YEARS, 1024 OTHERWISE
KB	KELLY-BOOTLE STANDARD UNIT	1012 BYTES	COMPROMISE BETWEEN 1000 AND 1024 BYTES
KiB	IMAGINARY KILOBYTE	1024 $\sqrt{-1}$ BYTES	USED IN QUANTUM COMPUTING
kb	INTEL KILOBYTE	1023.937528 BYTES	CALCULATED ON PENTIUM F.P.U.
Kb	DRIVEMAKER'S KILOBYTE	CURRENTLY 908 BYTES	SHRINKS BY 4 BYTES EACH YEAR FOR MARKETING REASONS
KBa	BAKER'S KILOBYTE	1152 BYTES	9 BITS TO THE BYTE SINCE YOU'RE SUCH A GOOD CUSTOMER

http://xkcd.com/394/

Application: Expression Trees

Any basic mathematical expression containing binary operators may be represented using a binary tree

For example, $3(4a + b + c) + d/5 + (6 - e)$



Application: Expression Trees

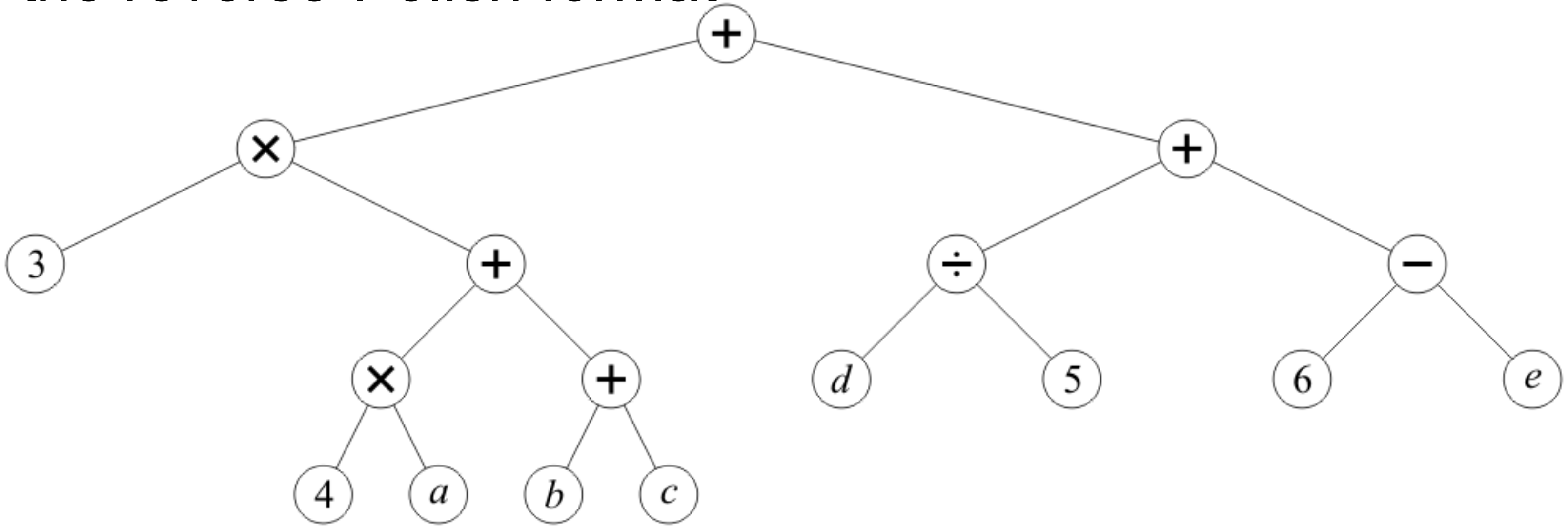
Observations:

- Internal nodes store operators
- Leaf nodes store literals or variables
- No nodes have just one sub tree
- The order is not relevant for
  - Addition and multiplication (commutative)
- Order is relevant for
  - Subtraction and division (non-commutative)
- It is possible to replace non-commutative operators using the unary negation and inversion:

$$(a/b) = a\,b^{-1} \qquad (a - b) = a + (-b)$$

Application: Expression Trees

A post-order depth-first traversal converts such a tree to the reverse-Polish format



$$3\ 4\ a\ \times\ b\ c\ +\ +\ \times\ d\ 5\ \div\ 6\ e\ -\ +\ +$$

Application: Expression Trees

Humans think in in-order

Computers think in post-order:

- Both operands must be loaded into registers
- The operation is then called on those registers

Most use in-order notation (C, C++, Java, C#, etc.)

- Necessary to translate in-order into post-order

Summary

In this talk, we introduced binary trees

- Each node has two distinct and identifiable sub-trees
- Either sub-tree may optionally be empty
- The sub-trees are ordered relative to the other

We looked at:

- Properties
- Applications