

Lecture 3: Algebraic Data Types

Søren Haagerup

Department of Mathematics and Computer Science
University of Southern Denmark, Odense

September 19, 2017

Algebraic Data Types

- Algebraic Data Types, are types which are constructed using *constructor* functions, and destructed using *pattern matching*
- Which Algebraic Data Types have we seen so far?
- *Bool, Maybe a, (a, b), [a]*
- Which non-algebraic types have we seen so far?
- *Int, Integer* etc., functions $a \rightarrow b$

The **data** keyword

If we wanted to define our own *Bool* data type, we would write

data *Bool* = *False* | *True*

This gives rise to constructor functions

- *False* :: *Bool*
- *True* :: *Bool*

corresponding patterns *False*, *True*.

Enumeration types in general

data *Bool* = *False* | *True*

data *Ordering* = *LT* | *EQ* | *GT*

data *Season* = *Spring* | *Summer* | *Fall* | *Winter*

data *MessageType* = *Success* | *Info* | *Warning* | *Danger*

data *ThreadState* = *Ready* | *Running* | *Terminated* | *Waiting*

If we have fixed number of elements in the type, enumeration types are useful.

isHot :: *Season* → *Bool*

isHot *Spring* = *True*

isHot *Summer* = *True*

isHot _ = *False*

Product types

- *Product types* correspond to structs/records in other languages. E.g

type *Age* = *Int*

type *Name* = *String*

data *Person* = *Person* *Age* *Name*

isOld :: *Person* → *Bool*

isOld (*Person* *a* _) | *a* ≥ 30 = *True*

isOld _ = *False*

- What type does the data constructor *Person* have?
- Why not just define a type synonym
type *Person* = (*Age*, *Name*) ?

Difference between tuples and general product types

- Let's say that we somewhere in a program have defined

```
type Age  = Int  
type Name = String  
type Person = (Age, Name)  
isOld :: Person → Bool
```

- And somewhere else, we have defined

```
type ExitCode = Int  
execCommand :: String → (ExitCode, String)
```

- Then we would be able to write

```
isOld (execCommand "ls")
```

which does not make much sense.

Product types

- When having written

data *Person* = *Person* *Age* *Name*

it is also reasonable to write functions

age :: *Person* → *Age*

age (*Person* *a* _) = *a*

name :: *Person* → *Name*

name (*Person* _ *n*) = *n*

Such functions are called “projection functions” or “accessor functions”

- They can be created by Haskell automatically by writing

data *Person* = *Person* {

age :: *Age*

name :: *Name*

}

Sum types

- We can also combine the concepts of an “Enumerated types” and “Product type”. We can define a type

data *Number* = *I Integer* | *D Double*

This makes it possible to store either an *Integer* or a *Double* in a *Number*, much like the JavaScript number type.

- Which constructors does it have, and what are the types of the constructors?

Recursive types

- Let's define a type

data $Nat = Z \mid S\ Nat$

- What does this type represent?
- What are the type signatures for the constructors?
- How can we implement $add :: Nat \rightarrow Nat \rightarrow Nat$?
- Come up with an alternative definition of add .
- How can we implement $toInt :: Nat \rightarrow Int$?
- Come up with an alternative definition of $toInt$.
- Which list function does the functions above remind us of?
- How can we implement $min :: Nat \rightarrow Nat \rightarrow Nat$?
- Which list function does this remind us of?

Example: File System

```
data Entry = File {  
    name :: String  
} | Folder {  
    name :: String,  
    entries :: [Entry]  
}  
  
entry :: Entry  
entry = Folder "Documents" [  
    File "File1",  
    File "File2",  
    Folder "Pictures" [  
        File "Picture 1",  
        File "Picture 2"  
    ]  
]
```

Bigger example: JSON data

data *Number* = *I Integer* | *D Double*

data *JsValue* = *JsArray* [*JsValue*]
 | *JsBoolean* *Bool*
 | *JsNull*
 | *JsNumber* *Number*
 | *JsObject* [(*String*, *JsValue*)]
 | *JsString* *String*

instance *Show Number* **where**

show (*I i*) = *show i*

show (*D d*) = *show d*

instance *Show JsValue* **where**

show (*JsArray xs*) = "[" ++ (concat (intersperse ", " (map show xs))) ++ "]"

show (*JsBoolean b*) = **if** *b* **then** "true" **else** "false"

show (*JsNull*) = "null"

show (*JsNumber n*) = *show n*

show (*JsObject xs*) = "{" ++ concat (intersperse ", "

["\"" ++ *k* ++ "\" : " ++ *show v* | (*k*, *v*) ← *xs*]) ++ "}"

show (*JsString s*) = *show s*

Polymorphic types

- Is the following a product or a sum type?

data *Pair a b = Pair a b*

- Another example of a polymorphic type

data *Maybe a = Nothing | Just a*

safeHead :: [a] → *Maybe a*

safeHead [] = *Nothing*

safeHead (x : xs) = *Just x*

- Another example of a polymorphic type

data *Either a b = Left a | Right b*

type *Error a = Either String a*

safeHead :: [a] → *Error a*

safeHead [] = *Left "Error: [] has no head!"*

safeHead (x : xs) = *Right x*

Error handling + Record update syntax

```
data Entry = File {name :: String}  
           | Folder {name :: String, entries :: [Entry]}
```

```
deriving Show
```

```
rename :: String → Entry → Entry
```

```
rename n e = e {name = n}
```

Polymorphic, recursive types

We could define a linked list as:

```
data List a = Nil | Cons a (List a)
```

The list type which Haskell includes has some syntactic sugar, so Haskell's built-in list corresponds to a definition like

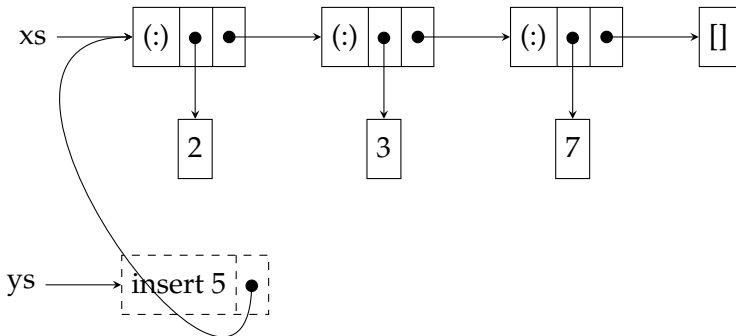
```
data [a] = [] | a : [a]  -- (not legal Haskell code!)
```

How to find the length of a list, defined using our own *List a* type?

```
length :: List a → Int
```

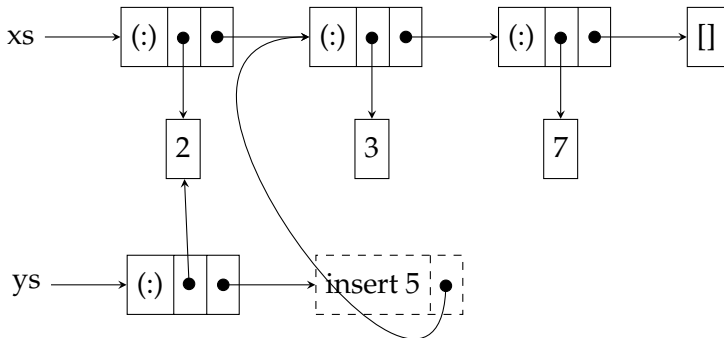
Insertion in a list

- Recall *insert x xs* which insert x at the correct position in a sorted list.
- We now want to investigate how the resulting list is defined:



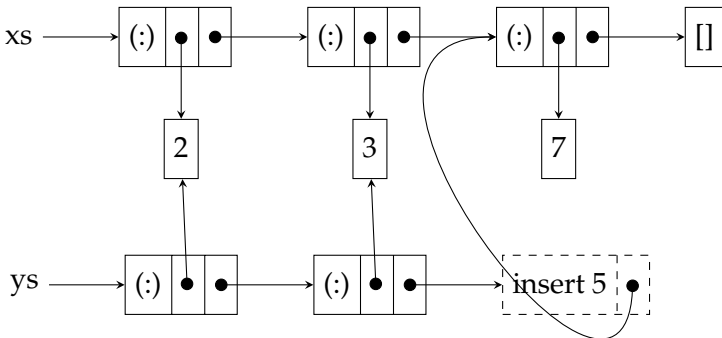
Insertion in a list

- Recall *insert x xs* which insert *x* at the correct position in a sorted list.
- We now want to investigate how the resulting list is defined:



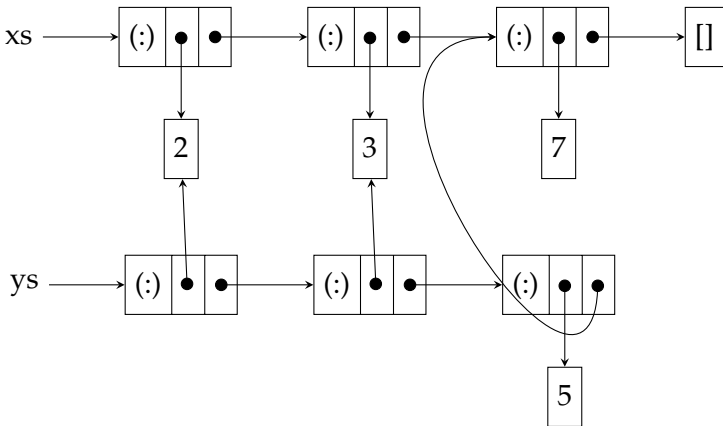
Insertion in a list

- Recall *insert x xs* which insert x at the correct position in a sorted list.
- We now want to investigate how the resulting list is defined:



Insertion in a list

- Recall *insert* x xs which insert x at the correct position in a sorted list.
- We now want to investigate how the resulting list is defined:



Immutable data structures

- never destroy data - only the garbage collector should do that!
- if we want to “change” an element, we generate a new one instead!

	Imperative	Functional
List	ArrayList	Linked List
Dictionary	HashMap	Balanced Tree

TREE STRUCTURES

A look at Data.Map

data Map k a

insert :: (Ord k) \Rightarrow k \rightarrow a \rightarrow Map k a \rightarrow Map k a -- O(log n)
lookup :: (Ord k) \Rightarrow k \rightarrow Map k a \rightarrow Maybe a -- O(log n)
delete :: (Ord k) \Rightarrow k \rightarrow Map k a \rightarrow Map k a -- O(log n)
update :: (Ord k) \Rightarrow
 (a \rightarrow Maybe a) \rightarrow k \rightarrow Map k a \rightarrow Map k a -- O(log n)
union :: (Ord k) \Rightarrow Map k a \rightarrow Map k a \rightarrow Map k a -- O(m + n)
member :: (Ord k) \Rightarrow k \rightarrow Map k a \rightarrow Bool -- O(log n)
size :: Map k a \rightarrow Int -- O(1)
map :: (a \rightarrow b) \rightarrow Map k a \rightarrow Map k b -- O(n)

How to use Data.Map

```

import Data.Map (Map)
import qualified Data.Map as Map

testMap = Map.fromList (zip [1..13] ['a'..'m'])

> putStrLn (Map.showTree testMap)
8 := 'h'
+ -- 4 := 'd'
| + -- 2 := 'b'
| | + -- 1 := 'a'
| | + -- 3 := 'c'
| + -- 6 := 'f'
|   + -- 5 := 'e'
|   + -- 7 := 'g'
+ -- 12 := 'l'
    + -- 10 := 'j'
    | + -- 9 := 'i'
    | + -- 11 := 'k'
    + -- 13 := 'm'
  
```

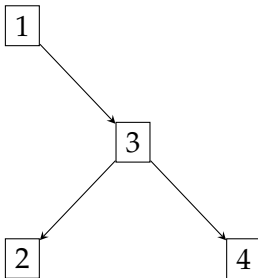
Binary Search Tree

data $Tree\ a = Nil \mid Node\ a\ (Tree\ a)\ (Tree\ a)$

Example:

$Node\ 1\ Nil\ (Node\ 3\ (Node\ 2\ Nil\ Nil)\ (Node\ 4\ Nil\ Nil))$

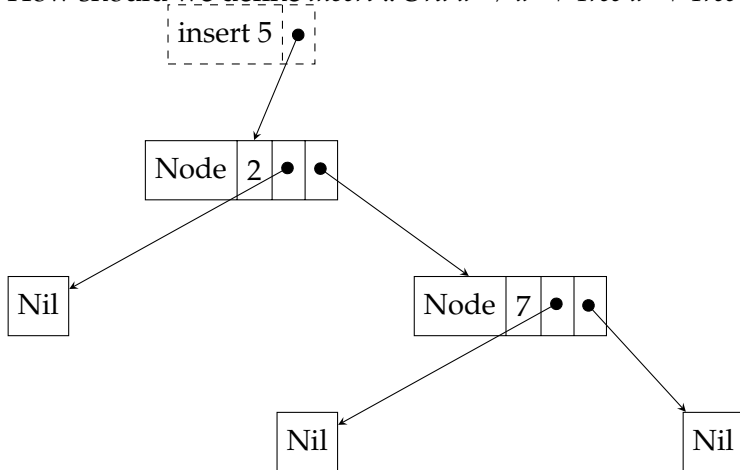
corresponds to the tree



Binary Search Tree

We are given **data** $Tree\ a = Nil \mid Node\ a\ (Tree\ a)\ (Tree\ a)$.

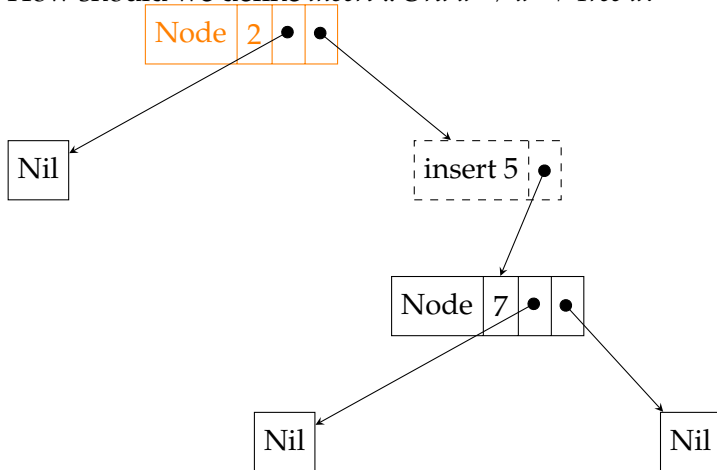
How should we define $insert :: Ord\ a \Rightarrow a \rightarrow Tree\ a \rightarrow Tree\ a$?



Binary Search Tree

We are given **data** $Tree\ a = Nil \mid Node\ a\ (Tree\ a)\ (Tree\ a)$.

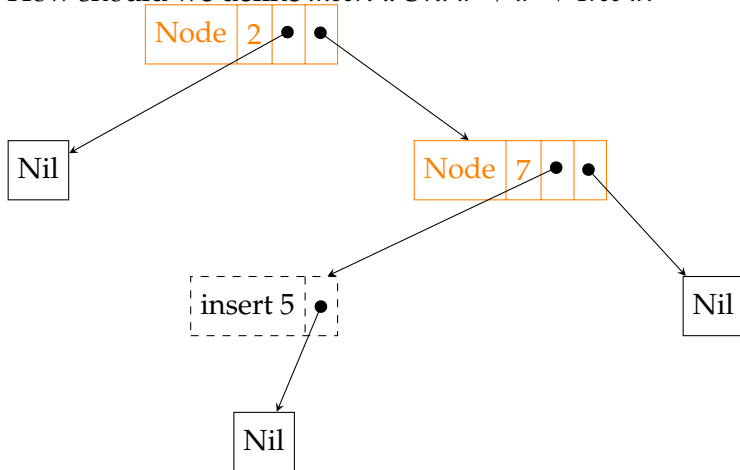
How should we define $insert :: Ord\ a \Rightarrow a \rightarrow Tree\ a$?



Binary Search Tree

We are given **data** $Tree\ a = Nil \mid Node\ a\ (Tree\ a)\ (Tree\ a)$.

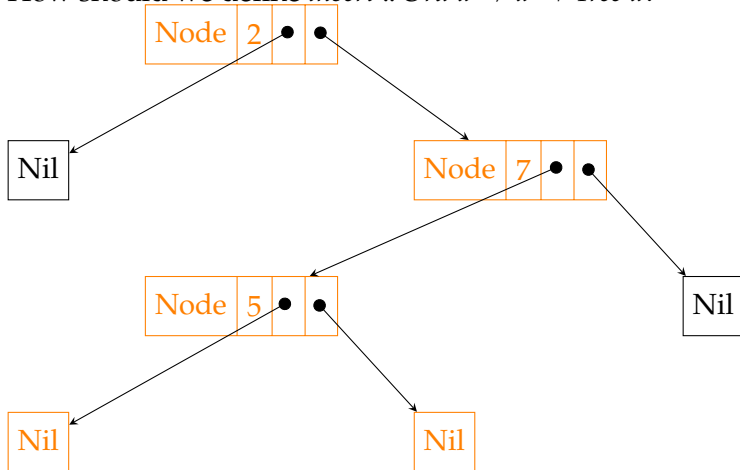
How should we define $insert :: Ord\ a \Rightarrow a \rightarrow Tree\ a$?



Binary Search Tree

We are given **data** $Tree\ a = Nil \mid Node\ a\ (Tree\ a)\ (Tree\ a)$.

How should we define $insert :: Ord\ a \Rightarrow a \rightarrow Tree\ a$?



insert - Binary search tree

insert :: Ord a \Rightarrow a \rightarrow Tree a \rightarrow Tree a

insert x Nil = Node x Nil Nil

insert x (Node y l r) = **case** (compare x y) **of**

LT \rightarrow Node y (insert x l) r

EQ \rightarrow Node x l r

GT \rightarrow Node y l (insert x r)

exists - Binary search tree

$exists :: Ord\ a \Rightarrow a \rightarrow Tree\ a \rightarrow Bool$

exists - Binary search tree

exists :: Ord a \Rightarrow a \rightarrow Tree a \rightarrow Bool

exists x Nil = False

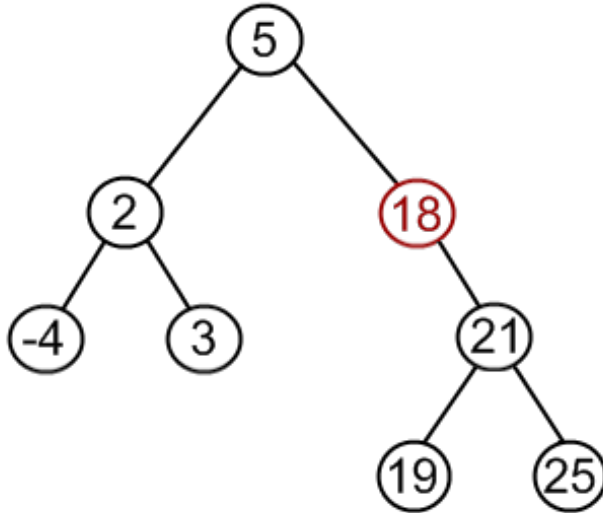
exists x (Node y l r) = **case** (compare x y) **of**

LT \rightarrow *exists* x l

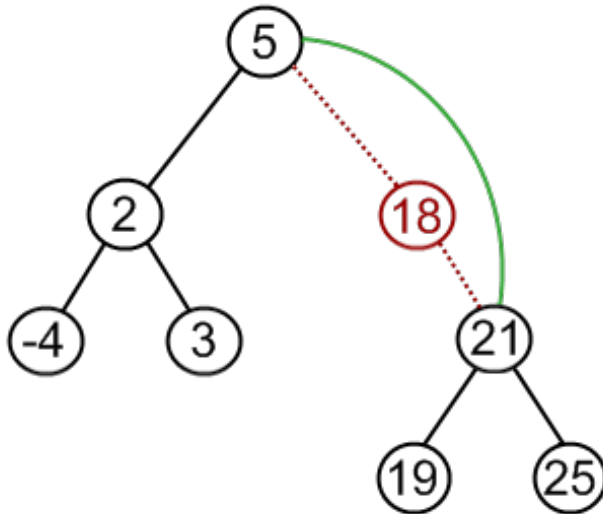
EQ \rightarrow True

GT \rightarrow *exists* x r

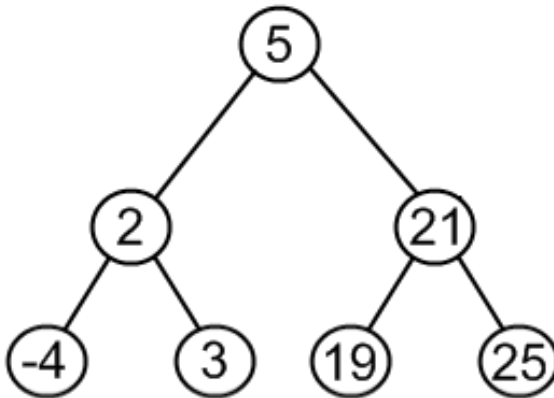
delete - Removal of node with 1 child



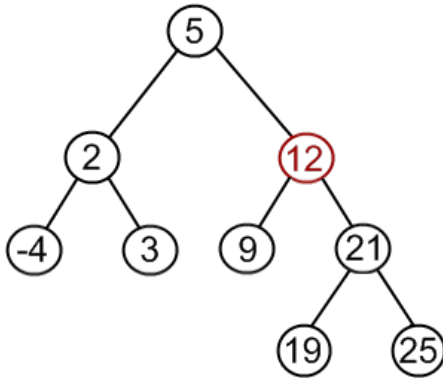
delete - Removal of node with 1 child



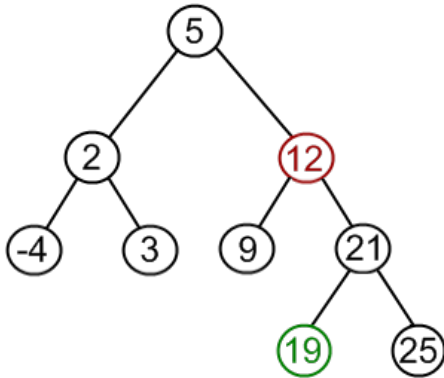
delete - Removal of node with 1 child



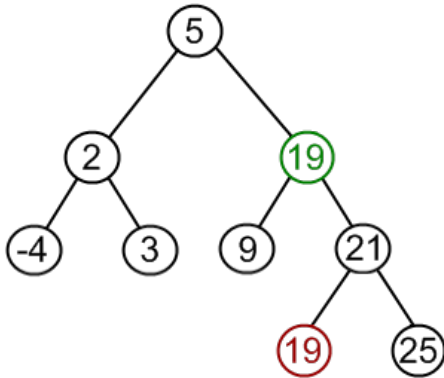
delete - Removal of node with 2 children



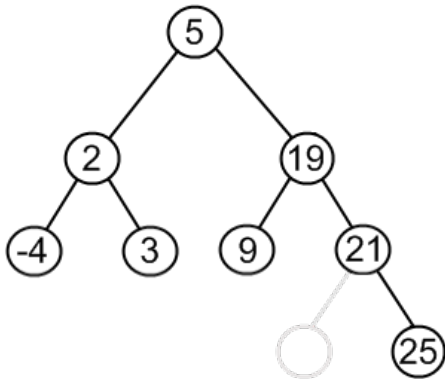
delete - Removal of node with 2 children



delete - Removal of node with 2 children



delete - Removal of node with 2 children



height and exists of tree - any binary tree

Now consider a binary tree without ordering constraints. How should we define these functions?

height :: *Tree a* → *Int*

exists :: *Eq a* ⇒ *a* → *Tree a* → *Bool*

height and exists of tree - any binary tree

height :: *Tree a* → *Int*

height Nil = 0

height (Node _ l r) = 1 + max (*height l*) (*height r*)

exists :: *Eq a* ⇒ *a* → *Tree a* → *Bool*

exists x Nil = *False*

exists x (Node y l r) = *x* ≡ *y* ∨ *exists x l* ∨ *exists x r*

Depth-first search - Inorder tree walk

We want to list all elements of the tree.

So for a binary search tree, the output will be a sorted list.

$$\textit{flatten} :: \textit{Tree } a \rightarrow [a]$$

Depth-first search - Inorder tree walk

flatten :: *Tree a* → [*a*]

flatten Nil = []

flatten (Node x l r) = *flatten l* ++ [*x*] ++ *flatten r*

Breath-first flatten

We want to output all nodes in the tree, but now as they would occur in a Breath-first search: All nodes from a level are returned from left to right, before the nodes from the next level.

$$bfflatten :: Tree\ a \rightarrow [a]$$

Breath-first flatten

$bfflatten\ t = bfs'\ [t]$

$bfs' :: [Tree\ a] \rightarrow [a]$

$bfs'\ [] = []$

$bfs'\ ts = [x \mid Node\ x\ _ _ \leftarrow ts] \mathbin{++} bfs'\ [t \mid Node\ _ \ l\ r \leftarrow ts, t \leftarrow [l, r]]$

Challenge: Depth first enumeration

Given any tree *Tree a*, output the tree *Tree Int* where each node should be numbered by its occurrence in a depth first search.

$$dfsnumber :: Tree\ a \rightarrow Tree\ Int$$

Challenge: Depth first enumeration

Given any tree *Tree a*, output the tree *Tree Int* where each node should be numbered by its occurrence in a depth first search.

$$dfsnumber :: Tree\ a \rightarrow Tree\ Int$$
$$dfsnumber\ t = snd\ (dfs\ (1, t))$$
$$dfs :: (Int, Tree\ a) \rightarrow (Int, Tree\ Int)$$
$$dfs\ (n, Nil) = (n, Nil)$$
$$dfs\ (n, Node\ v\ l\ r) = (n'', Node\ n\ l'\ r')$$

where

$$(n', l') = dfs\ (n + 1, l)$$
$$(n'', r') = dfs\ (n', r)$$

Red-Black Trees

data *Color* = *R* | *B*

data *Tree a* = *E* | *T Color (Tree a) a (Tree a)*

insert :: *Ord a* \Rightarrow *a* \rightarrow *Tree a* \rightarrow *Tree a*

insert x s = *makeBlack (ins s)*

where *ins E* = *T R E x E*

ins (T color a y b) | x < y = *balance color (ins a) y b*

| x \equiv y = *T color a y b*

| x > y = *balance color a y (ins b)*

makeBlack (T _ a y b) = *T B a y b*

Red-Black Trees

data $Color = R \mid B$

data $Tree\ a = E \mid T\ Color\ (Tree\ a)\ a\ (Tree\ a)$

$balance\ B\ (T\ R\ (T\ R\ a\ x\ b)\ y\ c)\ z\ d = T\ R\ (T\ B\ a\ x\ b)\ y\ (T\ B\ c\ z\ d)$

$balance\ B\ (T\ R\ a\ x\ (T\ R\ b\ y\ c))\ z\ d = T\ R\ (T\ B\ a\ x\ b)\ y\ (T\ B\ c\ z\ d)$

$balance\ B\ a\ x\ (T\ R\ (T\ R\ b\ y\ c)\ z\ d) = T\ R\ (T\ B\ a\ x\ b)\ y\ (T\ B\ c\ z\ d)$

$balance\ B\ a\ x\ (T\ R\ b\ y\ (T\ R\ c\ z\ d)) = T\ R\ (T\ B\ a\ x\ b)\ y\ (T\ B\ c\ z\ d)$

$balance\ color\ a\ x\ b = T\ color\ a\ x\ b$

Type classes for your ADT - *Eq, Ord, Show*

If the type a is an instance of the type class $Eq\ a$, we know that (\equiv) is defined for it.

```
class  $Eq\ a$  where  
   $(\equiv) :: a \rightarrow a \rightarrow Bool$ 
```

We can implement equality for **data** $Bool = False \mid True$ by

```
instance  $Eq\ (Bool)$  where  
   $True \equiv True = True$   
   $False \equiv False = True$   
   $\_ \equiv \_ = False$ 
```


Type classes for your ADT - *Eq*, *Ord*, *Show*

- This can be generalized further. Consider:

data *Tree a* = *Nil* | *Node a (Tree a) (Tree a)*

Assuming we can compare values of *a*, we can recursively define comparison for *Tree a* by

instance (*Eq a*) \Rightarrow *Eq (Tree a)* **where**

Nil \equiv *Nil* = *True*

(*Node v1 l1 r1*) \equiv (*Node v2 l2 r2*)

= *v1* \equiv *v2* \wedge *l1* \equiv *l2* \wedge *r1* \equiv *r2*

_ \equiv _ = *False*

- This function can be automatically derived by

data *Tree a* = *Nil* | *Node a (Tree a) (Tree a)* **deriving** *Eq*

- How would the “trivial” functions look like for *Ord* and *Show*?

Remember labs this Friday!

- 9 exercises about binary search trees
- You can apply many of the ideas we have seen today
- Ex. 8 and 9 about how to make it balanced