Praktische Informatik 3: Funktionale Programmierung
Vorlesung 8 vom 02.12.2014: Abstrakte Datentypen

Christoph Lüth

Universität Bremen

Wintersemester 2014/15

Organisatorisches

- Raumänderung nächste Woche:

Vorlesung	Di, 09.12.	12-14	NW2 A0242
Tutorium	Mi, 10.12.	08-10	SpT C4180
Tutorium	Mi, 10.12.	10-12	entfällt!
Tutorium	Mi, 10.12.	12-14	GW1 A0160
Tutorium	Mi, 10.12.	14-16	SFG 1020

  Die Tutorien am Donnerstag finden wie gewohnt statt.

- Grund ist eine internationale Tagung...

Fahrplan

- Teil I: Funktionale Programmierung im Kleinen
- Teil II: Funktionale Programmierung im Großen
  - Abstrakte Datentypen
  - Signaturen und Eigenschaften
  - Spezifikation und Beweis
- Teil III: Funktionale Programmierung im richtigen Leben

Inhalt

- Abstrakte Datentypen
  - Allgemeine Einführung
  - Realisierung in Haskell
  - Beispiele

Refakturierung im Einkaufsparadies

Warum Modularisierung?

- Übersichtlichkeit der Module

  Lesbarkeit

- Getrennte Übersetzung

  technische Handhabbarkeit

- Verkapselung

  konzeptionelle Handhabbarkeit

Abstrakte Datentypen

Definition (Abstrakter Datentyp)

Ein abstrakter Datentyp (ADT) besteht aus einem (oder mehreren) Typen und Operationen auf diesem.

- Werte des Typen können nur über die bereitgestellten Operationen erzeugt werden
- Eigenschaften von Werten des Typen werden nur über die bereitgestellten Operationen beobachtet
- Einhaltung von Invarianten über dem Typ kann garantiert werden

Implementation von ADTs in einer Programmiersprache:

- Benötigt Möglichkeit der Kapselung (Einschränkung der Sichtbarkeit)
- bspw. durch Module oder Objekte

ADTs vs. algebraische Datentypen

- Algebraische Datentypen
  - Frei erzeugt
  - Keine Einschränkungen
  - Insbesondere keine Gleichheiten
- ADTs:
  - Einschränkungen und Invarianten möglich
  - Gleichheiten möglich

ADTs in Haskell: Module

- Einschränkung der Sichtbarkeit durch Verkapselung
- Modul: Kleinste verkapselbare Einheit
- Ein Modul umfaßt:
  - Definitionen von Typen, Funktionen, Klassen
  - Deklaration der nach außen sichtbaren Definitionen
- Gleichzeitig: Modul $\hat{=}$ Übersetzungseinheit (getrennte Übersetzung)

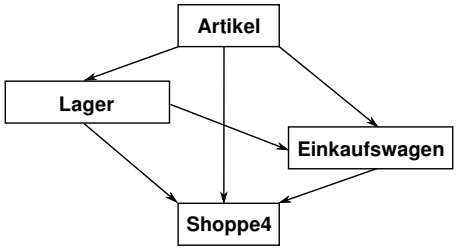
Module: Syntax

- Syntax:

```
module Name(Bezeichner) where Rumpf
```

- Bezeichner können leer sein (dann wird alles exportiert)
- Bezeichner sind:
  - Typen: T, T(c1, ..., cn), T(..)
  - Klassen: C, C(f1, ...,fn), C(..)
  - Andere Bezeichner: Werte, Felder, Klassenmethoden
  - Importierte Module: **module** M
- Typsynonyme und Klasseninstanzen bleiben sichtbar
- Module können rekursiv sein *(don't try at home)*

Refakturierung im Einkaufsparadies: Modularchitektur



Refakturierung im Einkaufsparadies I: Artikel

- Es wird alles exportiert
- Reine Datenmodellierung

```
module Artikel where

import Data.Maybe

-- Modellierung der Artikel.

data Apfel = Boskoop | CoxOrange | GrannySmith
             deriving (Eq, Show)
```


Refakturierung im Einkaufsparadies II: Lager

```
module Lager(
  Posten,
  artikel,
  menge,
  posten,
  cent,
  Lager,
  leeresLager,
  einlagern,
  suche,
  inventur
  ) where
```

- Implementiert ADTs Posten und Lager
- Garantierte Invarianten:
  - Posten hat immer die korrekte Artikel und Menge:

    ```
    posten :: Artikel→ Menge→ Maybe Posten
    ```
  - Lager enthält keine doppelten Artikel:

    ```
    einlagern :: Artikel→ Menge→ Lager→ Lager
    ```


Refakturierung im Einkaufsparadies III: Einkaufswagen

```
module Einkaufswagen(
  Einkaufswagen,
  leererWagen,
  einkauf,
  kasse,
  kassenbon
  ) where
```

- Implementiert ADT Einkaufswagen
- Garantierte Invariante:
  - Korrekte Artikel und Menge im Einkaufswagen

    ```
    einkauf :: Artikel→ Menge→
               Einkaufswagen→ Einkaufswagen
    ```
- Nutzt dazu Posten aus Modul Lager

Benutzung von ADTs

- Operationen und Typen müssen importiert werden
- Möglichkeiten des Imports:
  - Alles importieren
  - Nur bestimmte Operationen und Typen importieren
  - Bestimmte Typen und Operationen nicht importieren

Importe in Haskell

- Syntax:

```
import [qualified] M [as N] [hiding][(Bezeichner)]
```

- *Bezeichner* geben an, was importiert werden soll:
  - Ohne Bezeichner wird alles importiert
  - Mit **hiding** werden Bezeichner nicht importiert
- Für jeden exportierten Bezeichner f aus M wird importiert
  - f und qualifizierter Bezeichner M.f
  - **qualified**: nur qualifizierter Bezeichner M.f
  - Umbenennung bei Import mit as (dann N.f)
  - Klasseninstanzen und Typsynonyme werden immer importiert
- Alle Importe stehen immer am Anfang des Moduls

Beispiel

module A(x,y) **where**...

Import(e)	Bekannte Bezeichner
import A	x, y, A.x, A.y
import A()	*(nothing)*
import A(x)	x, A.x
import qualified A	A.x, A.y
import qualified A()	*(nothing)*
import qualified A(x)	A.x
import A hiding ()	x, y, A.x, A.y
import A hiding (x)	y, A.y
import qualified A hiding ()	A.x, A.y
import qualified A hiding (x)	A.y
import A as B	x, y, B.x, B.y
import A as B(x)	x, B.x
import qualified A as B	B.x, B.y

Quelle: Haskell98-Report, Sect. 5.3.4

Schnittstelle vs. Implementation

- Gleiche Schnittstelle kann unterschiedliche Implementationen haben
- Beispiel: (endliche) Abbildungen

Endliche Abbildungen

- Eine Sichtweise: Ersatz für Hashtables in imperativen Sprachen. Sehr nützlich!
- Abstrakter Datentyp für endliche Abbildungen:
  - Datentyp

    ```
    data Map α β
    ```

  - Leere Abbildung:

    ```
    empty :: Map α β
    ```

  - Abbildung auslesen:

    ```
    lookup :: Ord α⇒ α→ Map α β→ Maybe β
    ```

  - Abbildung ändern:

    ```
    insert :: Ord α⇒ α→ β→ Map α β→ Map α β
    ```

  - Abbildung löschen:

    ```
    delete :: Ord α⇒ α→ Map α β→ Map α β
    ```


Eine naheliegende Implementation

- Modellierung als Haskell-Funktion:

  ```
  data Map α β = Map (α→ Maybe β)
  ```

- Damit einfaches lookup, insert, delete:

  ```
  empty = Map (λx→ Nothing)

  lookup a (Map s) = s a

  insert a b (Map s) =
    Map (λx→ if x == a then Just b else s x)

  delete a (Map s) =
    Map (λx→ if x == a then Nothing else s x)
  ```

- Instanzen von Eq, Show nicht möglich
- Speicherleck

Endliche Abbildungen: Anwendungsbeispiel

- Artikel im Lager:

  ```
  data Lager = Lager (M.Map Artikel Posten)
  ```

- Artikel suchen:

  ```
  suche a (Lager l) = fmap menge (M.lookup a l)
  ```

- Ins Lager hinzufügen:

  ```
  einlagern :: Artikel→ Menge→ Lager→ Lager
  einlagern a m (Lager l) =
    case posten a m of
      Nothing → Lager l
      Just p → Lager (M.insert a p l)
  ```

- Für Inventur fehlt Möglichkeit zur Iteration
- Daher: Map als Assoziativliste

Map als Assoziativliste

```
newtype Map α β = Map [(α, β)]
```

- Zusatzfunktionalität:
  - Iteration (fold)

    ```
    fold :: Ord α⇒ ((α, β)→ γ→ γ)→ γ→ Map α β→ γ
    fold f e (Map ms) = foldr f e ms
    ```

  - Instanzen von Eq und Show

    ```
    instance (Eq α, Eq β)⇒ Eq (Map α β) where
      Map s1 == Map s2 =
        null (s1 \\ s2) && null (s1 \\ s2)
    ```

  - ... ist aber ineffizient (Zugriff/Löschen in $\mathcal{O}(n)$)
- Deshalb: balancierte Bäume

AVL-Bäume und Balancierte Bäume

AVL-Bäume

Ein Baum ist ausgeglichen, wenn

- alle Unterbäume ausgeglichen sind, und
- der Höhenunterschied zwischen zwei Unterbäumen höchstens eins beträgt.

Balancierte Bäume

Ein Baum ist balanciert, wenn

- alle Unterbäume balanciert sind, und
- für den linken und rechten Unterbaum l, r gilt:

$$size(l) \leq w \cdot size(r) \quad (1)$$
$$size(r) \leq w \cdot size(l) \quad (2)$$

w — Gewichtung (Parameter des Algorithmus)

Implementation von balancierten Bäumen

- Der Datentyp

  ```
  data Tree α = Null
              | Node Weight (Tree α) α (Tree α)
  ```

- Gewichtung (Parameter des Algorithmus):

  ```
  type Weight = Int

  weight :: Weight
  ```

- Hilfskonstruktor, setzt Größe (l, r balanciert)

  ```
  node :: Tree α→ α→ Tree α→ Tree α
  node l n r = Node h l n r where
                 h = 1+ size l+ size r
  ```

- Hilfskonstruktor, balanciert ggf. neu aus:

  ```
  mkNode :: Tree α→ α→ Tree α→ Tree α
  ```

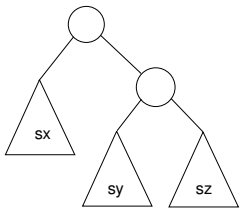
Balance sicherstellen

- Problem:

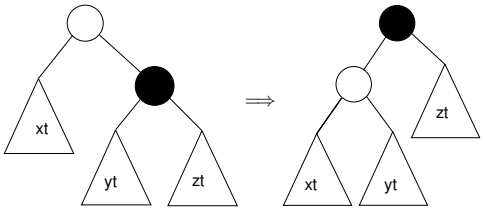
  Nach Löschen oder Einfügen zu großes Ungewicht

- Lösung:

  Rotieren der Unterbäume

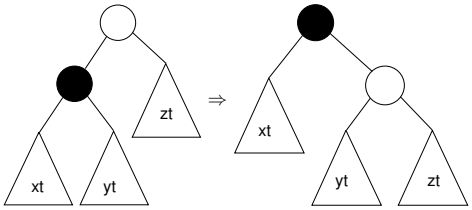


Linksrotation




```
rotl :: Tree α→ Tree α
rotl (Node _ xt y (Node _ yt x zt)) =
  node (node xt y yt) x zt
```

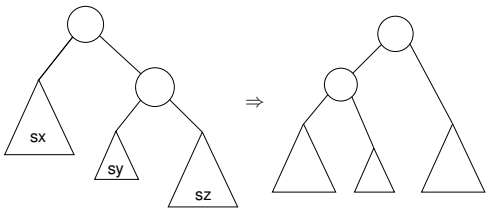

Rechtsrotation




```
rotr :: Tree α→ Tree α
rotr (Node _ (Node _ ut y vt) x rt) =
  node ut y (node vt x rt)
```

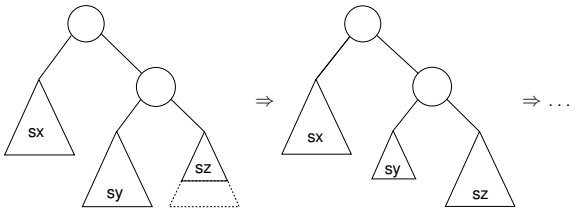

Balanciertheit sicherstellen

- Fall 1: Äußerer Unterbaum zu groß
- Lösung: Linksrotation



Balanciertheit sicherstellen

- Fall 2: Innerer Unterbaum zu groß oder gleich groß
- Reduktion auf vorherigen Fall durch Rechtsrotation des Unterbaumes



Balance sicherstellen

- Hilfsfunktion: Balance eines Baumes

```
bias :: Tree α→ Ordering
bias Null = EQ
bias (Node _ lt _ rt) = compare (size lt) (size rt)
```

- Zu implementieren: mkNode lt y rt
  - Voraussetzung: lt, rt balanciert
  - Konstruiert neuen balancierten Baum mit Knoten y
- Fallunterscheidung:
  - rt zu groß, zwei Unterfälle:
    - Linker Unterbaum von rt kleiner (Fall 1): bias rt == *LT*
    - Linker Unterbaum von rt größer/gleich groß (Fall 2): bias rt == *EQ*, bias rt == *GT*
  - lt zu groß, zwei Unterfälle (symmetrisch).

Konstruktion eines ausgeglichenen Baumes

- Voraussetzung: lt, rt balanciert

```
mkNode lt x rt
  | ls+ rs < 2  = node lt x rt
  | weight* ls < rs  =
      if bias rt == LT then rotl (node lt x rt)
      else rotl (node lt x (rotr rt))
  | ls > weight* rs =
      if bias lt == GT then rotr (node lt x rt)
      else rotr (node (rotl lt) x rt)
  | otherwise   = node lt x rt where
    ls = size lt; rs= size rt
```


Balancierte Bäume als Maps

- Endliche Abbildung: Bäume mit (key, value) Paaren

```
type Map α β = Tree (α, β)
```

- insert fügt neues Element ein:

```
insert :: Ord α⇒ α→ β→ Map α β→ Map α β
insert k v Null = node Null (k, v) Null
insert k v (Node n l a@(kn, _) r)
  | k < kn  = mkNode (insert k v l) a r
  | k == kn = Node n l (k, v) r
  | k > kn  = mkNode l a (insert k v r)
```

- lookup liest Element aus
- remove löscht ein Element
  - Benötigt Hilfsfunktion join :: Tree α→ Tree α→ Tree α

Zusammenfassung Balancierte Bäume

- Auslesen, einfügen und löschen: logarithmischer Aufwand ($\mathcal{O}(\log n)$)
- Fold: linearer Aufwand ($\mathcal{O}(n)$)
- Guten durchschnittlichen Aufwand
- Auch in der Haskell-Bücherei: Data.Map (mit vielen weiteren Funktionen)

Defizite von Haskells Modulsystem

- Signatur ist nur implizit
  - Exportliste enthält nur Bezeichner
  - Wünschenswert: Signatur an der Exportliste annotierbar, oder Signaturen in separater Datei
  - In Java: Interfaces
- Keine parametrisierten Module
  - Vgl. Lager
  - In ML-Notation:

```
module Lager(Map : MapSig) : LagerSig =...

module Lager1 = Lager(MapList)
module Lager2 = Lager(MapFun)
```

  - In Java: abstrakte Klassen

ADTs vs. Objekte

- ADTs (Haskell): Typ plus Operationen
- Objekte (z.B. Java): Interface, Methoden.
- Gemeinsamkeiten:
  - Verkapselung (information hiding) der Implementation
- Unterschiede:
  - Objekte haben internen Zustand, ADTs sind referentiell transparent;
  - Objekte haben Konstruktoren, ADTs nicht (Konstruktoren nicht unterscheidbar)
  - Vererbungsstruktur auf Objekten (Verfeinerung für ADTs)
  - Java: `interface` eigenes Sprachkonstrukt
  - Java: `packages` für Sichtbarkeit

Zusammenfassung

- Abstrakte Datentypen (ADTs):
  - Besteht aus Typen und Operationen darauf
- Realisierung in Haskell durch Module
- Beispieldatentypen: endliche Abbildungen
- Nächste Vorlesung: ADTs durch Eigenschaften spezifizieren