Grundlagen der Informatik 1
WS 2008/09

Prof. Mühlhäuser, Dr. Rößling, Melanie Hartmann, Daniel Schreiber
http://proffs.tk.informatik.tu-darmstadt.de/gdi1

Übung 3 v1.1 03.11.2008

Änderungen

- 08.11.2008: Aufgabe 5.1.3 Wurzel eines binären Codierungsbaums als zweiten Parameter für `encode-list` hinzugefügt.

1 Mini Quiz

Kreuzen Sie die wahren Aussagen an!

☐ Strukturdefinitionen können mit `local` innerhalb von Funktionen erfolgen.

☐ Eine Baumstruktur in Scheme kann verschiedene Datentypen in ihren Knoten speichern.

☐ Innerhalb eines `local` Ausdrucks kann nicht auf Definitionen außerhalb des `local` Ausdruck zugegriffen werden.

☐ Der Scope einer Namensbindung ist der textuelle Bereich, in dem sich ein Auftreten des Namens auf diese Namensbindung bezieht.

2 Fragen

1. Welche Richtlinien sollten bei der Benutzung von `local` beachtet werden?
2. Was macht aus Software-Engineering Sicht hochwertigen Code aus?
3. Erläutere Probleme, die entstehen können, wenn Daten redundant abgelegt werden.

3 λocalverbot

Hintergrund: RGB Der RGB Code eines Farbton ist ein Tripel $(R, G, B) R, G, B \in [0; 255]$. Die Werte R, G und B sind dabei die Anteile der Grundfarben Rot, Grün und Blau am Farbton. Sie reichen von 0 (diese Grundfarbe ist gar nicht im Farbton vorhanden) bis 255 (diese Farbe ist in voller Intensität im Farbton vorhanden). Beispiel: (255 0 0) reines Rot, (114 247 160) Hellgrün. Der RGB Code eignet sich gut, um Farben für die Darstellung auf einem Bildschirm zu speichern. Der vom Menschen empfundene Farbton lässt sich aus einem RGB Tripel nur schwer ersehen. Was für

eine Farbe istz.B. (204, 102, 153)?. Um dieses Problem zu lösen, kann man den Winkel H (H für hue/Farbe) zu einem RGB Tripel berechnen. Ein H in der Nähe von 0° bedeutet z.B. rötlich, ein H in der Nähe von 240° bedeutet bläulich. Aus Wikipedia stammen folgende Formeln zur Errechnung des Winkels H zu einem RGB Tripel (R, G, B):

$$r = R/255, g = G/255, b = B/255$$

$$M = max(r, g, b), m = min(r, g, b)$$

$$H = \begin{cases} 0° & \text{if } M = m \ , \\ 60°(g-b)/(M-m) & \text{if } r = M \ , \\ 120° + 60°(b-r)/(M-m) & \text{if } g = M \ , \\ 240° + 60°(r-g)/(M-m) & \text{if } b = M \end{cases}$$

Scheme verwendet die Struktur `color` zum speichern von RGB Tripeln. Folgende Scheme Prozedur berechnet H.

```
;; Diese Funktion berechnet aus einer RBG Farbe den hue Wert
;; http://de.wikipedia.org/wiki/HSI-Farbmodell
;; http://de.wikipedia.org/wiki/RGB-Farbraum
(define (hue color)
  (local (
    (define r (/ (color-red color) 255))
    (define b (/ (color-blue color) 255))
    (define g (/ (color-green color) 255))
    (define (h x)
      (if (< x 0) (+ x 360) x)))
    (local (
        (define MAXIMUM (max r b g))
        (define MINIMUM (min r b g)))
        (local (
           (define (f a b)
             (* (/ (- a b) (- MAXIMUM MINIMUM)) 60)))
             (cond
                 [(= MINIMUM MAXIMUM) 0]
                 [(= r MAXIMUM) (h (f g b)) ]
                 [(= g MAXIMUM) (h (+ (f b r) 120))]
                 [else (h (+ (f r g) 240)) ])))))
```

Auftrag Der Software Entwickler „Schematics“ gibt Ihnen den Auftrag die Prozedur `hue` so umzuändern, so dass sie ohne die Verwendung von `local` Ausdrücken auskommt, da diese der Firmenpolitik widersprechen. Hilfsprozeduren sollen auch nicht verwendet werden, da sie nur unnötigen Dokumentationsaufwand verursachen. Ersetzen Sie eine `local` Definition nach der anderen. Was halten Sie von der Firmenpolitik von „Schematics“?

4 Tree Sort

Diese Aufgabe hilft bei der Lösung der Hausübung!
In dieser Aufgabe werden Sie den TreeSort Algorithmus zum sortieren von Listen implementieren. Dazu wird eine in der Informatik sehr oft verwendete Datenstruktur, der Binärbaum benötigt.

Hintergrund: Die Datenstruktur sortierter Binärbaum

Ein Baum ist eine rekursive Datenstruktur, die folgendermaßen definiert ist: Jeder *Baumknoten* enthält einen Inhalt, einen linken Sohn und einen rechten Sohn. Linker und rechter Sohn sind dabei auch wieder *Baumknoten*. Als Rekursionsanker dient der „leere Baum“, `empty`, der per Definition ein Baumknoten ist. Sind beide Söhne eines Baumknotens `empty`, so nennt man diesen Baumknoten ein „Blatt“. Der oberste Baumknoten, der selber nicht Sohn eines weiteren Baumknoten ist heißt „Wurzel“ des Baumes.
Zusätzlich gilt bei einem sortierten binären Baum folgende Bedingung: Sei n ein Bauknoten. Der Inhalt des linken Sohns von n ist ein sortierter Binärbaum, dessen größtes Blatt stets kleiner oder gleich dem Inhalt von n oder der linke Sohn ist der „leere Baum“; der Inhalt des rechten Sohns von n ist ein sortierter Binärbaum, dessen kleinstes Blatt stets größer als der Inhalt von n oder der rechte Sohn von n ist der „leere Baum“. Vergleichen Sie dazu folgende Scheme Definition und die Beispiele:

```
1  ;; struct for storing binary trees
2  (define-struct treenode (left content right))
3
4  ;;example binary tree with three nodes
5  ;;     2
6  ;;    / \
7  ;;   1   3
8  ;;  / \ / \
9  ;;
10 (make-treenode (make-treenode empty 1 empty) 2 (make-treenode empty 3
     empty))
11
12 ;;not a sorted binary tree! Left son is larger than the node content!
13 ;;     2
14 ;;    / \
15 ;;   4   3
16 ;;  / \ / \
17 ;;
18 (make-treenode (make-treenode empty 4 empty) 2 (make-treenode empty 3
     empty))
```

4.1 Sortieren von Zahlen

Lesen Sie zunächst aufmerksam alle Teilaufgaben durch bevor sie mit der Implementierung der Lösung beginnen. Entscheiden Sie sich für ein Vorgehen top-down oder bottom-up und machen sich eine Wunschliste von Funktionen. Verwenden Sie `local` für Funktionen die nicht nach außen sichtbar sein sollen.

1. Implementieren Sie eine Funktion `sort-list`, die alle Zahlen in einer Liste zuerst in einen sortierten Binärbaum einfügt und dann als sortierte Liste wieder zurück gibt.

2. Implementieren Sie eine Funktion `tree-insert-list`, die eine Liste von Zahlen in einen sortierten Binärbaum einfügt.

3. (K) Implementieren Sie eine Funktion `tree-insert`, die die Wurzel $root$ eines sortierten Binärbaums T und eine Zahl n übergeben bekommt und die die Wurzel eines neuen sortierten Binärbaum T' zurück liefert, der alle Elemente von T sowie n enthält. Beachten Sie dabei die Regeln für sortierte binäre Bäume, nachdem der linke Sohn immer kleiner und der rechte Sohn immer größer als der Inhalt des Vaters sein muss!

4.2 Sortieren von Studenten

Implementieren Sie nun Varianten von `sort-list`, um damit Studenten nach verschiedenen Kriterien sortieren zu können.

1. Überlegen Sie sich vorab eine Struktur zur Speicherung von Studenten mit Namen und Geburtsdatum.

2. Erstellen sie zwei Varianten von `sort-list`, die Listen von Studenten sortieren.
   a) Sortieren Sie die Studenten nach ihren Matrikelnummern
   b) Sortieren Sie die Studenten nach ihrem Alter

Hausübung

Die Vorlagen für die Bearbeitung werden im GdI1-Portal bereitgestellt. Kommentieren Sie Ihren selbst erstellten Code. Die Hausübung muss bis zum Abgabedatum im GdI1-Portal abgegeben werden. Der Fachbereich Informatik misst der Einhaltung der Grundregeln der wissenschaftlichen Ethik großen Wert bei. Zu diesen gehört auch die strikte Verfolgung von Plagiarismus. Mit der Abgabe Ihrer Hausübung bestätigen Sie, dass Sie bzw. Ihre Gruppe alleiniger Autor des gesamten Materials sind. Falls Ihnen die Verwendung von Fremdmaterial gestattet war, so müssen Sie dessen Quellen deutlich zitieren.

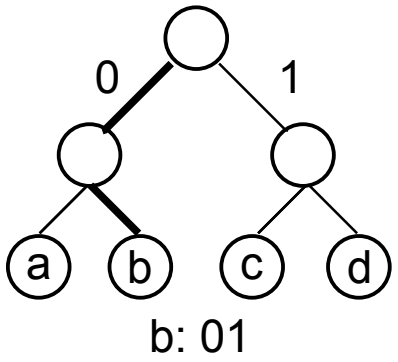
Abgabe: Spätestens Fr, 14.11.08, 16:00. Wichtig: Um die volle Punktzahl zu erlangen, müssen Sie jede Ihrer Prozeduren und Hilfsprozeduren mindestens mit Vertrag, Beschreibung und Beispiel kommentieren, sowie jeweils Testfälle angeben. Verwenden Sie als Sprachlevel für die gesamte Hausübung „Zwischenstufe“.

5 Datenkompression

Codierung

Codierung ist die Darstellung von Symbolen in einem Alphabet durch Folgen von Nullen und Einsen. Eine bekannte Codierung ist z.B. ASCII. Im ASCII Code werden alle Symbole durch eine Folge von acht Nullen und Einsen codiert. Ein anderer bekannter Code ist der Morse-Code. Hier wird das e mit der kurzen Folge 0 (. = 0) codiert, das seltene q hingegen mit 1101 (- - . -; 1 = -). Werden seltene Symbole durch lange Codes und häufige Symbole durch kurze Codes dargestellt, sinkt die Gesamtlänge der Codesequenz für Nachrichten die aus vielen Symbolen bestehen. Diese Methode wird in Datenkompressionsverfahren, wie z.B. dem ZIP Algorithmus verwendet.
Als Basis der Codierung dienen binäre Codierungsbäume. In den Blättern dieser Bäume stehen die zu codierenden Symbole. Möchte man den Code zu einem Symbol ermitteln, so beginnt man bei

der Wurzel des Baumes und durchläuft den Baum bis zum Blatt in dem das Symbol steht. Geht man dabei nach links, so fügt man eine „0“ an den Code an, geht man nach rechts, so fügt man eine „1“ an.



5.1 Encode (5P)(K)

In der Vorlage ist ein einfacher Codierungsbaum `simplecode` enthalten. Benutzen Sie diesen zum Testen ihrer Prozeduren.

1. Schreiben Sie eine Funktion `contains?`, die überprüft, ob ein Symbol als Blatt in einem binären Codierungsbaum vorkommt. Verwenden Sie Rekursion über die Baumstruktur: Ein Symbol kommt in einem binären Codierungsbaum vor, wenn es im Wurzelknoten steht oder im linken oder rechten Sohn vorkommt.

2. Schreiben Sie eine Funktion `encode`, die ein Symbol und die Wurzel eines binären Codierungsbaums übergeben bekommt. Als Ergebnis soll die Codierung des Symbols mit dem Codierungsbaum zurück geliefert werden. Verwenden Sie Rekursion über die Baumstruktur, d.h. wenn das Symbol in der Wurzel vorkommt wird der bisher erzeugte Code zurück geliefert. Kommt das Symbol im linken Teilbaum vor, so fügen Sie eine „0“ in den Code ein und rufen `encode` rekursiv mit dem linken Sohn auf. Kommt das Symbol im rechten Teilbaum vor, so fügen Sie eine „1“ in den Code ein und rufen `encode` rekursiv mit dem rechten Sohn auf.

3. Schreiben Sie eine Funktion `encode-list`, die eine Liste von Symbolen und die Wurzel eines binären Codierungsbaums übergeben bekommt und eine Liste mit den Codes der Symbole zurück liefert.

5.2 Komprimierende Codes (8P)

Der Codierungsbaum `simplecode` erzeugt zu allen Symbolen gleichlange Codes. Um Daten zu komprimieren sollte man aber, wie beim Morse Code, für häufig vorkommende Symbole kurze Codes wählen und dafür für seltene Symbole längere Codes erlauben. Um die Codelänge optimal auf die Symbolhäufigkeit abzustimmen, werden Wahrscheinlichkeitsbäume als Codebäume verwendet. In einem binären Wahrscheinlichkeitsbaum enthält der Inhalt jedes Knotens eine Wahrscheinlichkeit, die gleich der Summe der Wahrscheinlichkeiten des linken und rechten Sohnes ist. Die Struktur `ptree-nodes` aus der Vorlage soll für den Knoteninhalt von Wahrscheinlichkeitsbäumen verwendet werden.
Die Liste `freq` von `ptree-nodes` aus der Vorlage enthält die Wahrscheinlichkeit des Auftretens von Buchstaben in einem englischen Text. Die Vorlage enthält ausserdem die Funktion `make-subtree`,

die aus einem `ptree-node` einen einelementigen Teilbaum erzeugt. Bekommt `make-subtree` einen Baum übergeben, so bleibt dieser unverändert.

1. Implementieren Sie die Funktion `make-subtree-list`, die `make-subtree` auf alle Elemente einer Liste anwendet. Die Elemente der Liste sind entweder `ptree-nodes` oder vom Typ `treenode`, wobei der Inhalt der Wurzel wiederum vom Typ `ptree-node` ist.

2. Schreiben Sie die Funktion `sort-ptree-list`, die eine Liste von Wahrscheinlichkeitsbäumen, jeweils repräsentiert durch ihren Wurzelknoten, nach der Wahrscheinlichkeit sortiert. Orientieren Sie sich an der Lösung der Präsenzübun: Statt Studenten nach der Matrikelnummer, sind Wahrscheinlichkeitsbäumen nach der Wahrscheinlichkeit Ihres Wurzelknotens zu sortieren. Sortieren Sie `(make-subtree-list freq)`.

3. Schreiben Sie eine Funktion `build-ptree`. Die Funktion erhält eine Liste von noch zu bearbeitenden Wahrscheinlichkeitsbäumen, repräsentiert durch ihre Wurzelknoten, als Parameter und soll folgendes tun:
   1. Suchen Sie aus allen unbearbeiteten Wahrscheinlichkeitsbäumen, die beiden mit der geringsten Wahrscheinlichkeit im Wurzelknoten aus. Die beiden Wurzelknoten nennen wir t_1 und t_2, wobei die Wahrscheinlichkeit von t_1 kleiner gleich der von t_2 ist.
   2. Löschen Sie t_1 und t_2 aus der Liste der unbearbeiteten Wahrscheinlichkeitsbäumen und fügen Sie einen neuen Wurzelknoten t_3 hinzu. Der linke Sohn von t_3 ist t_1, der rechte ist t_2. Die Wahrscheinlichkeit für t_3 ist die Summe der Wahrscheinlichkeiten von t_1 und t_2. Als Symbol setzen Sie bei t_3 `'unused` ein.
   3. Wiederholen Sie diese Schritte, bis Sie nur noch einen Wahrscheinlichkeitsbaum in der Liste der unbearbeiteten Wahrscheinlichkeitsbäumen finden und liefern Sie diesen zurück.

4. Ändern Sie die Funktion `encode` zu `ptree-encode`, `contains?` zu `ptree-contains?` und `encode-list` zu `ptree-encode-list`, so dass diese mit Wahrscheinlichkeitsbäumen als Codierungsbäumen umgehen können.

5. Codieren Sie die Symbolfolge „informatik macht spass“ mit der `ptree-encode-list` Funktion und dem mit `build-ptree` erzeugten Codierungsbaum. Die Funktion `build-ptree-list` muss dazu mit der Liste `(make-subtree-list freq)` als Parameter aufgerufen werden.