

Kapitel 6

Praktische Datenstrukturen

Eine besonders praktische Datenstruktur ist der *Binärbaum*. Auf Binärbäumen lassen sich eine Reihe praktischer Datenstrukturen aufbauen. In diesem Kapitel dienen Suchbäume und Bäume für Huffman-Codierung als Beispiele.

6.1 letrec

Es ist Zeit für ein weiteres Sprachelement: Für Rekursion war es bisher häufig notwendig, eine separate Hilfsprozedur zu schreiben, welche die rekursiven Aufrufe vornimmt. Ein typisches Beispiel ist die endrekursive Fassung von `factorial`:

```
(define factorial
  (lambda (n)
    (factorial-1 n 1)))

(define factorial-1
  (lambda (n result)
    (if (= n 1)
        result
        (factorial-1 (- n 1) (* n result)))))
```

Dies ist auf Dauer etwas umständlich, da jedesmal ein neuer Name verbraucht wird. Außerdem müssen ggf. zusätzliche Daten an die Hilfsprozedur übergeben werden müssen, welche bei jedem rekursiven Aufruf unverändert durchgereicht werden.

Die offensichtliche Abhilfe wäre, die Hilfsprozedur an eine lokale Variable zu binden. Hierzu steht uns `let` zur Verfügung. Leider erlaubt `let` keine Rekursion, da die Namen auf der linken Seite in den Ausdrücken auf der rechten Seite nicht sichtbar sind. Ebenso ist es bei `let*` (schließlich auch nur syntaktischer Zucker für `let`), bei dem zusätzlich zur `let`-Situation auch noch die darüberliegenden Bindungen sichtbar werden. Darum gibt es noch ein drittes Mitglied der `let`-Familie, mit Syntax identisch zu der der beiden anderen. `Letrec` zeichnet sich dadurch aus, daß alle Variablen auf den linken Seite in den Ausdrücken auf den rechten Seiten sichtbar sind.

```
(define factorial
  (lambda (n)
    (letrec
      ((factorial-1
        (lambda (n result)
          (if (= n 1)
              result
              (factorial-1 (- n 1) (* n result)))))))
```

```
(factorial-1 (- n 1) (* n result))))))
(factorial-1 n 1)))
```

Letrec ist nach derzeitigem Kenntnisstand noch nicht in andere, einfachere Sprachelemente übersetzbar.

6.2 Binäre Bäume

Binäre Bäume treten in verschiedenen Varianten auf. Die hier betrachteten binären Bäume (meist kurz „Bäume“ genannt) gehören zu einer der folgenden Klassen:

- ein *leerer Baum*,
- ein *Blatt*, das ganz unten in einem Baum steht und eine *Markierung* besitzt,
- ein (innerer) *Knoten*, der sich innen im Baum befindet, und
 - eine *Markierung*,
 - einen *linken Teilbaum* und
 - einen *rechten Teilbaum*

besitzt.

(In vielen Realisierungen von Bäumen fehlen entweder die leeren Bäume oder die Knoten. Blätter lassen sich beispielsweise als Knoten mit zwei leeren Teilbäumen repräsentieren. Der Vollständigkeit halber gibt es hier alle drei.)

Bäume lassen sich mit dem üblichen Schema repräsentieren und abstrahieren. Dabei halten drei unterschiedliche Typen für die drei unterschiedlichen Sorte Bäume her. Zunächst die leeren Bäume:

```
(define empty-tree-type (make-type 'empty-tree))

(define make-empty-tree
  (lambda ()
    (make-typed-object empty-tree-type #f)))

(define empty-tree? (typed-object-predicate empty-tree-type))
```

Das #f kommuniziert dabei keine Information, aber aufgrund der Schnittstelle von `make-typed-object` muß irgendetwas dort stehen.

Blätter führen noch die Markierung mit:

```
(define leaf-type (make-type 'leaf))

(define make-leaf
  (lambda (label)
    (make-typed-object leaf-type label)))

(define leaf? (typed-object-predicate leaf-type))

(define leaf-label
  (lambda (leaf)
    (typed-object-value leaf-type leaf)))
```

Knoten sind repräsentiert aus einem Paar, dessen `car` die Markierung und dessen `cdr` ein Paar aus linkem und rechtem Teilbaum ist:

```

(define node-type (make-type 'node))

(define make-node
  (lambda (label left-branch right-branch)
    (make-typed-object node-type
      (cons label
        (cons left-branch
          right-branch)))))

(define node? (typed-object-predicate node-type))

(define node-label
  (lambda (node)
    (car (typed-object-value node-type node))))

(define node-left-branch
  (lambda (node)
    (car (cdr (typed-object-value node-type node)))))

(define node-right-branch
  (lambda (node)
    (cdr (cdr (typed-object-value node-type node)))))

```

Basierend auf diesen Prozeduren lassen sich zwei gelegentlich nützliche abgeleitete Prozeduren definieren.

```

(define nonempty-tree-label
  (lambda (tree)
    (if (leaf? tree)
        (leaf-label tree)
        (node-label tree))))

(define tree?
  (lambda (value)
    (or (empty-tree? value)
        (leaf? value)
        (node? value))))

```

6.3 Suchbäume

Eine typische Anwendung für Bäume sind *Suchbäume*, eine Repräsentation für Mengen. Suchbäume arbeiten effizienter als Listen, wenn sie Mengen repräsentieren: Eine Liste muß Element für Element abgearbeitet werden um festzustellen, ob sich ein Element darin befindet. Bei Suchbäumen muß auf den Elementen eine totale Ordnung definiert sein. Suchbäume sind dann Bäume mit folgender Eigenschaft:

- Alle Markierungen des linken Teilbaums eines Knotens haben eine kleinere Markierung als die Markierung des Knotens.
- Alle Markierungen des rechten Teilbaums eines Knotens haben eine größere Markierung als die Markierung des Knotens.

Also läßt sich durch Ansicht der Markierung eines Knotens feststellen (wenn diese nicht sowieso schon die gesuchte ist), in welchem Teilbaum des Knotens eine gesuchte Markierung stecken muß.

Für Suchbäume wird eine ein neuer Typ definiert:

```
(define search-tree-type (make-type 'search-tree))
```

Suchbäume bestehen — neben dem Baum selbst — noch aus der Ordnung und einer Gleichheitsrelation (die natürlich zueinander passen müssen):

```
(define make-search-tree
  (lambda (= < tree)
    (make-typed-object search-tree-type
      (cons (cons = <) tree))))
```

Dabei `make-search-tree` nur „intern“ verwendet werden. Dieser Konstruktor ist in gewisser Weise gefährlich, da er es erlaubt, Werte vom Typ mit Namen `search-tree` zu erzeugen, welche die Suchbaumeigenschaft nicht erfüllen. Stattdessen fangen alle Suchbäume beim leeren Suchbaum an:

```
(define make-empty-search-tree
  (lambda (= <)
    (make-search-tree = < (make-empty-tree))))
```

Prädikat und Selektoren gehen streng schematisch:

```
(define search-tree? (typed-object-predicate search-tree-type))
```

```
(define search-tree-<
  (lambda (search-tree)
    (cdr (car (typed-object-value search-tree-type search-tree)))))
```

```
(define search-tree-=
  (lambda (search-tree)
    (car (car (typed-object-value search-tree-type search-tree)))))
```

```
(define search-tree-tree
  (lambda (search-tree)
    (cdr (typed-object-value search-tree-type search-tree))))
```

Zunächst zum Elementtest. Dazu werden die Komponenten des Suchbaums extrahiert; dann wird eine Hilfsprozedur aufgerufen, die nicht jedesmal alles auspacken muß:

```
(define search-tree-member?
  (lambda (element search-tree)
    (let ((= (search-tree-= search-tree))
          (< (search-tree-< search-tree)))
      (letrec
        ((member?
          (lambda (tree)
            (cond
              ((empty-tree? tree)
               #f)
              ((= (nonempty-tree-label tree) element)
               #t)
              ((leaf? tree)
               #f)
              ((< element (node-label tree))
               (member? (node-left-branch tree)))
              (else
               (member? (node-right-branch tree))))))))
      (member? (search-tree-tree search-tree)))))
```

Die Prozedur `search-tree-insert` fügt ein neues Element in den Suchbaum ein und erhält dabei die Suchbaumeigenschaft.

```
(define search-tree-insert
  (lambda (element search-tree)
    (let ((= (search-tree= search-tree))
          (< (search-tree-< search-tree)))
      (letrec
        ((insert
          (lambda (tree)
            (cond
              ((empty-tree? tree)
               (make-leaf element))
              ((= element (nonempty-tree-label tree))
               tree)
              ((leaf? tree)
               (make-node element
                           (if (< element (leaf-label tree))
                               (make-empty-tree)
                               tree)
                           (if (< element (leaf-label tree))
                               tree
                               (make-empty-tree))))))
          ((< element (node-label tree))
           (make-node (node-label tree)
                     (insert (node-left-branch tree))
                     (node-right-branch tree)))
          (else
           (make-node (node-label tree)
                     (node-left-branch tree)
                     (insert (node-right-branch tree)))))))
      (make-search-tree = < (insert (search-tree-tree search-tree))))))
```

6.4 Huffman-Bäume

Eine andere der vielen Anwendungen von Binärbäumen liegt in der Codierungstheorie: *Huffman-Bäume* dienen dazu, Daten als Folge von Bits (also von Nullen und Einsen) zu repräsentieren. Für Textdaten gibt es z.B. eine ganze Reihe von Standards, in denen jeweils ein Buchstabe durch eine feste Folge von Bits dargestellt wird. Beispiele hierfür sind ASCII (7 Bits, für den angelsächsischen Sprachraum) und ISO Latin 1 (8 Bits, für den westeuropäischen Sprachraum). Diese Codierungen werden auch *Codierungen mit fester Länge* genannt.

Die Codierungen fester Länge haben den Nachteil, daß Symbole, die so gut wie gar nicht in realen Texten vorkommen, trotzdem den gleichen Platz in der Codierung einnehmen wie solche, die ständig vorkommen. Deswegen ist es gelegentlich vorteilhaft, Codierungen mit variabler Länge zu verwenden, in denen häufig verwendete Symbole weniger Bits beanspruchen als selten verwendete Symbole. (Im Morse-Alphabet ist „E“ ein einzelner Punkt.)

Codierungen variabler Länge sind schwieriger zu konstruieren als solche mit fester Länge: Angenommen, „E“ ist codiert als 0, und „S“ ist codiert als 00. Steht nun die Bitfolge 000 für „EEE“, „SE“ oder „ES“? Eine Lösung für das Problem ist, einen speziellen Separator-Code zu verwenden, um die Grenzen zwischen Symbolen zu markieren. (Eine Pause im Morse-Alphabet.) Eine andere Lösung ist es, die Symbole so zu codieren, daß niemals der Code für ein Symbol der Anfang (oder

Präfix) des Codes eines anderen Zeichens ist. Solche Codierungen heißen *Präfix-Codierungen*. Der Code für „E“ und „S“ oben erfüllt diese Eigenschaft nicht, da 0 ein Präfix von 00 ist.

Effektive Präfix-Codierungen benutzen kurze Codes für häufig vorkommende und lange für selten vorkommende Symbole. Eine Methode für die Konstruktion solcher Codierungen ist die *Huffman-Methode*. Eine Huffman-Codierung kann als binärer Baum repräsentiert werden, dessen Blätter mit den Symbolen markiert sind. Jeder Knoten ist mit der Menge der Symbole der Blätter markiert, die unter ihm liegen. Zusätzlich wird jedes Blatt und jeder Knoten mit einem Gewicht markiert: Bei Blättern steht das Gewicht für die relative Häufigkeit des Symbols, bei einem Knoten für die Summe der Gewichte unter ihm. Die Gewichte werden weder für die Codierung noch für die Decodierung benötigt, werden aber bei der Konstruktion eines Huffman-Baums nützlich sein.

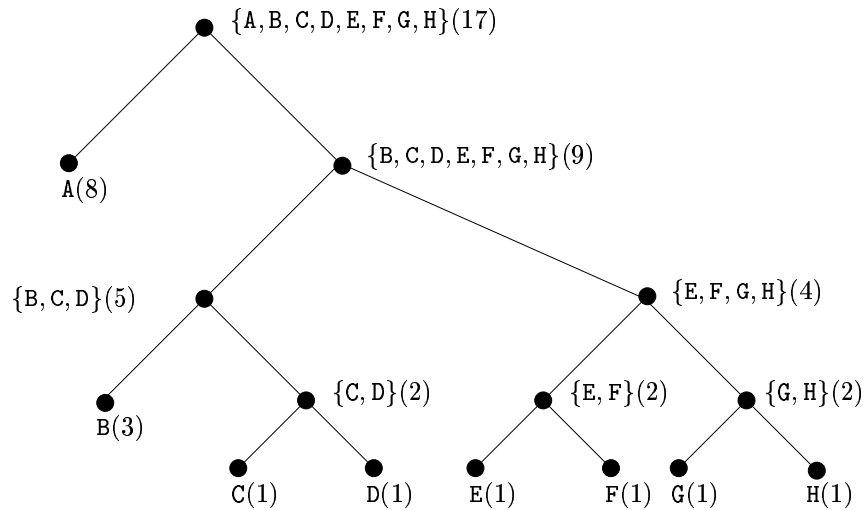


Abbildung 6.1: Ein Huffman-Baum

Abbildung 6.1 zeigt einen Huffman-Baum mit den Buchstaben A–H als Symbole, wobei A die relative Häufigkeit 8, B die relative Häufigkeit 3 und alle anderen Symbole die relative Häufigkeit 1 haben. Die Codierung eines Symbols ergibt sich aus dem Pfad von der Wurzel des Baums bis zu dem Blatt, welches das Symbol als Codierung hat: Bei jeder Abzweigung nach links wird eine 0 an die Codierung gehängt, bei jeder Abzweigung nach rechts eine 1. Für D zum Beispiel ist die Codierung 1011.

Die Decodierung funktioniert entsprechend: Die Bitfolge kennzeichnet einen Pfad innerhalb des Baums, dabei steht eine 0 für eine Abzweigung nach links, eine 1 für eine Abzweigung nach rechts. Ist ein Blatt erreicht, so ist das entsprechende Symbol erkannt und ein neues beginnt.

Für die Repräsentation von Huffman-Bäumen dienen ganz normale binäre Bäume, ohne zusätzliche Typ-Unterscheidung. Blätter sind mit einem Symbol einem Gewicht markiert:

```
(define make-huffman-leaf
  (lambda (symbol weight)
    (make-leaf (cons symbol weight))))
```

```
(define huffman-leaf-symbol
  (lambda (leaf)
```



```

      (let ((next-branch
            (choose-branch (car bits) current-branch)))
        (if (leaf? next-branch)
            (cons (huffman-leaf-symbol next-branch)
                  (decode-1 (cdr bits) tree))
            (decode-1 (cdr bits) next-branch))))))
    (decode-1 bits tree)))

(define choose-branch
  (lambda (bit branch)
    (cond
      ((= bit 0) (node-left-branch branch))
      ((= bit 1) (node-right-branch branch))
      (else (error "bad bit")))))

```

Die Codierung findet Symbol für Symbol statt; die entstehenden Bitfolgen werden aneinandergelängt:

```

(define huffman-encode
  (lambda (message tree)
    (fold-right (lambda (symbol rest)
                  (append (huffman-encode-symbol symbol tree)
                          rest))
                '()
                message)))

```

Die Codierung muß für ein Symbol das Blatt finden, das damit markiert ist. Da sich der Weg zum passenden Blatt nicht von der Wurzel aus sehen läßt, muß der Baum durchsucht werden. `huffman-encode-symbol` liefert `#f`, wenn das Symbol `symbol` nicht im Baum enthalten ist, und sonst eine Bitfolge, die das Symbol codiert:

```

(define huffman-encode-symbol
  (lambda (symbol tree)
    (cond
      ((node? tree)
       (let ((maybe-encoding
             (huffman-encode-symbol symbol (node-left-branch tree))))
         (if maybe-encoding
             (cons 0 maybe-encoding)
             (let ((maybe-encoding
                   (huffman-encode-symbol symbol (node-right-branch tree))))
               (if maybe-encoding
                   (cons 1 maybe-encoding)
                   #f))))))
      ((equal? (huffman-leaf-symbol tree)
                symbol)
       '())
      (else
       #f))))

```

Es verbleibt nun doch Problem, zu einer gegebenen Aufstellung der relativen Häufigkeiten der Symbole einer Sprache den entsprechenden Huffman-Baum zu konstruieren. Grundlegendes Prinzip ist dabei, den Baum so zu konstruieren, daß die Symbole mit der geringsten relativen Häufigkeit möglichst weit unten im Baum platziert werden. („Weit unten“ heißt ja gerade „lange Codierung“.)

Begonnen wird mit einer Menge der Blätter des Baums, konstruiert aus der Tabelle der Symbole und ihren relativen Häufigkeiten. Nun werden zwei Elemente mit geringstmöglichen Häufigkeiten aus der Menge entfernt, zu einem Binärbaum mit gerade diesen Elementen als linken und rechten Teilbaum verschmolzen, und der neu gewonnen Baum der Menge wieder hinzugefügt. Dieser Vorgang wird sukzessive wiederholt, bis die Menge einelementig ist; sie enthält dann gerade den Huffman-Baum für die Tabelle. Der entsprechende Prozess für den Beispielbaum sieht folgendermaßen aus:

Anfang	$\{(A(8)), (B(3)), (C(1)), (D(1)), (E(1)), (F(1)), (G(1)), (H(1))\}$
Verschmelzen	$\{(A(8)), (B(3)), (\{C, D\}(2)), (E(1)), (F(1)), (G(1)), (H(1))\}$
Verschmelzen	$\{(A(8)), (B(3)), (\{C, D\}(2)), (\{E, F\}(2)), (G(1)), (H(1))\}$
Verschmelzen	$\{(A(8)), (B(3)), (\{C, D\}(2)), (\{E, F\}(2)), (\{G, H\}(2)), \}$
Verschmelzen	$\{(A(8)), (B(3)), (\{C, D\}(2)), (\{E, F, G, H\}(4))\}$
Verschmelzen	$\{(A(8)), (\{B, C, D\}(5)), (\{E, F, G, H\}(4))\}$
Verschmelzen	$\{(A(8)), (\{B, C, D, E, F, G, H\}(9))\}$
Verschmelzen	$\{(\{A, B, C, D, E, F, G, H\}(17))\}$