

Informatik I

Einfache Datenstrukturen

G. Zachmann
 Clausthal University, Germany
zach@in.tu-clausthal.de

Motivation

- **Datenstrukturen** sind höheres Organisationskonzept
 - Vgl. die bislang behandelten Binärcodierungen elementarer Datentypen
- Algorithmen & Datenstrukturen sind 2 Seiten derselben Medaille!
- In diesem Abschnitt **Grundbausteine von Datenstrukturen**
 - Einige höhere Datenstrukturen werden später behandelt werden
- Ein „klassisches Buch“



Niklaus Wirth: *Algorithmen und Datenstrukturen*; Pascal Version, 5. Auflage, Teubner, 1999.

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 2

Array



- Ein **eindimensionales Array** besteht aus einer bestimmten Anzahl von Datenelementen
 - Elemente haben gleichen Typ → **homogenes** Array (C, Java, allg. eher in statisch typisierten Sprachen)
 - Verschiedenen Typ → **inhomogenes** Array (Python, Smalltalk, ..., allg. eher in dynamisch typisierten Sprachen)
- Beispiel: Vektor, Zeile oder Spalte einer Tabelle
 - Z.B. Abtastung eines Signals zu konstanten Zeitintervallen

Zeitpunkt	1	2	3	4	...	30	31
Signalmärkte	10.5	10.5	12.2	9.8	...	13.1	13.3
- Elemente werden indiziert, d.h., Identifikation und Zugriff erfolgt über **Index** = ganze Zahl ≥ 0 (typ. der Form **a[i]**)
- Auf jedes Element des Array kann mit demselben, **konstanten Zeitaufwand** zugegriffen werden

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 3

Mehrdimensionale Arrays

- **Zweidimensionale Arrays** speichern die Werte mehrerer eindimensionaler Zeilen in Tabellen-(Matrix-)Form
 - Syntax: `a[i][j]`
- Analog *n*-dimensionale Arrays
- **Array von Arrays**
 - ist auch 2-dim. Datenstruktur
 - Nicht notw. quadratisch
 - In den meisten Sprachen anders zu erzeugen / zuzugreifen / implementiert als (quadratisches) 2-dim. Array
- In Python gibt es eigtl. nur letzteres; in C++ gibt es beides

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 4

Mathematische Interpretation

- Array = Funktion $A : \mathbb{N} \mapsto T$, $T = \text{Typ des Arrays (= der Elemente)}$
- Beispiel: eine Funktion $t : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \mapsto \mathbb{R}$, die einem Koordinatentripel einen Temperaturwert zuordnet (Wettersimulation)
 - Wert der Funktion an der Stelle $(1,1,3)$, also $t(1,1,3)$, findet sich dann in $t[1][2][3]$
- Arrays eignen sich in der Praxis grundsätzlich nur dann zur Speicherung einer Funktion, wenn diese **dicht** ist, d.h., wenn die Abbildung für die allermeisten Indexwerte definiert ist
 - Sonst würde eine Arraydarstellung viel zuviel Platz beanspruchen
 - Außerdem geht dies nur für endliche Funktionen
- Wichtiger Spezialfall : strings = array of char
 - Viele Programmiersprachen haben dafür eigene Syntax / Implementierung

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 5

Zeit-Aufwand für elementare Operationen

- Annahme: Array enthält N Elemente
- Element Nr i lesen: konstant [O(1)]
- Element an Position i einfügen: $\sim N$ [O(N)]
- Element Nr i löschen: $\sim N$ [O(N)]
- Array löschen: konstant [O(1)]

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 6

Records, Structs, {Klassen} (Verbunde)

- Oft bestehen aber auch **Beziehungen zwischen Werten unterschiedlichen Typs**
 - Etwa zwischen Name und Monatsverdienst eines Beschäftigten
- Wir verbinden zusammengehörige Daten unterschiedlichen Typs zu einem **Verbund = record, struct, Klasse**
- Einzelteile eines Records / Structs / Klasse heißen **Attribute** oder **Members**
- Beispiel: Stammdaten

Name	"Meistermann"
Vorname	"Martin"
GebTag	10
GebMonat	05
GebJahr	1930
Familienstand	"verheiratet"
...	...

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 7

- Übliche Syntax zur Auswahl: Punkt-Notation
 - Beispiel: `s.name` oder `s.birthday`
 - Manchmal auch Pfeil-Notation: `s->name` oder `s->birthday`
- Komponenten eines Verbunds können von beliebigem Typ sein
- Also auch wieder Verbunde, Arrays, etc.
- Seien T_1, \dots, T_n die Typen der Members, dann hat der Record/Struct den (algebraischen) Typ $T_1 \times \dots \times T_n$

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 8

Verkettete Strukturen (linked structures)

"The name of the song is called 'Haddock's Eyes.' "

"Oh, that's the name of the song, is it?" Alice said, trying to feel interested.


"No, you don't understand," the Knight said, looking a little vexed. "That's what the name is called. The name really is 'The Aged Aged Man.' "

"Then I ought to have said 'That's what the song is called'?" Alice corrected herself.

"No, you oughtn't: that's quite another thing! The song is called 'Ways and Means,' but that is only what it's called, you know!"

"Well, what is the song, then?" said Alice, who was by this time completely bewildered.

"I was coming to that," the Knight said. "The song really is 'A-sitting On A Gate,' and the tune's my own invention."



Lewis Carroll
Through the Looking Glass

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 9

Verkettete vs. Sequentielle Allokation (Allocation)

- Ziel: Menge von Objekten abarbeiten
- **Sequential allocation:** ein Objekt nach dem anderen anordnen
 - Maschinenebene: aufeinanderfolgende Speicherstellen
 - Python / C++: Array von Objekten
- **Linked allocation:** jedes Objekt enthält Link / Zeiger / Referenz auf das nächste
 - Maschinenebene: Zeiger ist Speicheradresse des nächsten Objektes
 - Python: `object1.next = object2` ("alles ist ein Zeiger")
- Hauptunterschied:
 - Sequentiell: Indizierung wird unterstützt
 - Verkettet: Vergrößerung und Verkleinerung ist einfach
- Achtung: in Python gibt es scheinbar(!) beides für umsonst

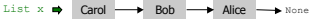
G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 10

Verkettete Liste (Linked List)

- Liste = Folge von Elementen a_0, a_1, \dots, a_{n-1}
 - Elemente sind geordnet: a_i ist Nachfolger von a_{i-1} (wie bei Array)
 - es können an beliebiger Stelle Elemente eingefügt und wieder entfernt werden (i.A. anders als bei Array)
- Implementierung:


```
class List:
    def __init__( self ):
        self.name = ""
        self.next = None
```

 - Üblicherweise mit Hilfe von verketteten Listenelementen
 - Listenelement enthält
 - "Nutzdaten" (satellite data) = eigentliche Elemente a_i
 - Zeiger auf nachfolgendes Listenelement



G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 11

Verkettete Liste Demo

```
a = List()
a.name = "Alice"
a.next = None
b = List()
b.name = "Bob"
b.next = a
c = List()
c.name = "Carol"
c.next = b
```

addr	value
c0	0
c1	0
c2	0
c3	0
c4	0
c5	0
c6	0
c7	0
c8	0
c9	0
CA	0
CB	0

main memory

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 12

Traversierung einer Liste

- Musterbeispiel für das traversieren einer mit null endenden verketteten Liste

```

l = List()
... Liste füllen ...
li = l
while li != None:
    print li.name
    li = li.next
    
```

```

graph LR
    x((x)) --> Carol[Carol]
    Carol --> Bob[Bob]
    Bob --> Alice[Alice]
    Alice --> None[None]
    
```

```

$ ./list.py
    
```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 13

Liste mit mehr innerem "Wissen"

- Anforderungen:
 - Anhängen soll in 1 Schritt gehen → Liste muß letztes Element (*tail*) kennen
 - Am Anfang einfügen auch → Liste muß Anfang (*head*) kennen
 - Methode um "nächstes" Element zu erfragen (*Iterator*) → "*Cursor*" verwalten
- Liste soll Interna kapseln (verstecken):
 - Elemente der Liste verstecken
 - Head und Tail speichern
 - Cursor verwalten

```

graph LR
    x((List x)) --> head[x.head]
    head --> Carol[Carol]
    Carol --> Bob[Bob]
    Bob --> Alice[Alice]
    Alice --> None[None]
    x --> tail[x.tail]
    tail --> Alice
    x --> cursor[x.CURSOR]
    cursor --> Bob
    
```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 14

```

class List:
    class ListElement:
        def __init__( self ):
            self.item = self.next = None

    def __init__( self ):
        self.head = None
        self.tail = None
        self.cursor = None

    def isEmpty(self):
        return self.head == None
    
```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 15

```

... (cont'd)
def append(self, item):
    if self.isEmpty():
        self.cursor = self.head = \
            self.tail = ListElement()
    else:
        self.tail.next = ListElement()
        self.tail = self.tail.next
        self.tail.item = item
        self.tail.next = None
    
```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 16

```

# methods dealing with the iterator (cursor)

def rewind(self):
    self.cursor = head

def getCurrentItem(self):
    if self.cursor == None:          # Spezialfall abfangen!
        return None
    return self.cursor.item

def getNextItem(self):
    if self.cursor == None:
        return None
    self.cursor = self.cursor.next
    return getCurrentItem()         # nicht etwa Code wiederholen!

```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 17

```

def insertAfterCurrent(self, item):
    if self.isEmpty():
        self.append(item)
        return
    if self.cursor == None:
        return # eigentlich nicht so gut
    z = ListElement()
    z.item = item
    z.next = self.cursor.next
    self.cursor.next = z

```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 18

```

def getNode( self, index ):
    z = self.head
    while index > 0 and z.next:
        z = z.next
        index -= 1
    return z

def insert( self, node, index ):
    ...

def findNode( self, item ):
    ...

```

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 19

Weitere Operationen

- `getCursorPos()`: Position (Index) des aktuellen Elementes
- `setCursorAtPos(i)`: Setze aktuelles Element auf den Index i
- `delete()`: lösche ganze Liste
- `removeCurrent()`: lösche aktuelles Element aus Liste
- `insertBeforeCurrent(item)`: Setzt Element e vor die aktuelle Position; Achtung: Aufwand im worst-case ~ N
- `find(item)`:
 - Suche **item** und setze Cursor auf entsprechende Zelle
 - Aufwand im worst-case ~ N

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 20

removeCurrent():

- Entfernt Element an aktueller Position
- Cursor zeigt anschließend auf nächstes Element (falls vorhanden, sonst auf Head)
- Achtung: Aufwand kann proportional zu N sein (Man muß erst das Element vor aktueller Position finden)

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 21

Eigenschaften der einfach verketteten Liste

- man kann schnell auf Elemente hinter der aktuellen Position zugreifen
- will man auf Elemente davor zugreifen, muß man immer beim Anfang der Liste beginnen und die Position suchen,
 - Problem z.B. bei `removeCurrent()`, `insertBeforeCurrent()`
- Asymmetrie im Aufwand beim Durchlaufen der Kette

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 22

Doppelt verkettete Liste

- Lösung: Doppelt verkettete Liste (*doubly linked list*)
- verkettet die Elemente in beide Richtungen
- Symmetrie im Aufwand beim Durchlaufen der Kette
- größerer Speicheraufwand
- Größerer Aufwand bei Entfernen / Einfügen

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 23

Multi-Listen

- Auch *mehrdimensionale* Listen genannt
- Menge von Elementen gleichzeitig nach mehreren Kriterien organisiert
- Beispiel: Liste aller Studenten, mit Teilliste aller Informatik-Studenten
- Ziel: Elemente nur 1x vorhalten, aber verschiedene Listen / Teillisten
- Lösung: jede Organisation durch eine Verkettung dargestellt

- Jede Liste kann für sich getrennt verwaltet werden

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 24

Beispiel: Dünnbesetzte Matrizen (*sparse matrix*)

- Matrix heißt **dünn besetzt**, wenn nur "wenige" Elemente $\neq 0$ sind
 - "Wenig" ist Definitionssache, z.B. 10%
- Multi-Liste ist gängige Methode, um dünnbesetzte Matrix zu implementieren

The diagram illustrates a multi-list implementation of a sparse matrix. It consists of a grid of nodes. The top row of nodes is labeled 'anker 11', 'anker 12', 'anker 13', and 'anker 14'. The left column of nodes is labeled 'anker 21', 'anker 22', 'anker 23', 'anker 24', and 'anker 21'. Each node is a small square containing a small circle. Arrows point from each node to its right and down neighbors, forming a grid of arrows. The non-zero elements of the matrix are represented by small squares containing a plus sign (+) at the intersections of the arrows. The non-zero elements are located at (1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4), and (5,1).

G. Zachmann Informatik 1 - WS 05/06 Datenstrukturen 25