Algorithmen und Datenstrukturen 1

Felix Ichters, Lukas Dzielski*

Sommersemester 2023

Begleitmaterial zur Vorlesung 'Algorithmen und Datenstrukturen'.

*Universität Heidelberg

Inhaltsverzeichnis

1 Beweise **4**
- 1.1 Statements . . . 4
- 1.2 Formal mathematical proofs . . . 4
- 1.3 Set definition rule: Beispiel . . . 5
- 1.4 Macro-steps in proofs . . . 5
- 1.5 Einfache Beweistechniken . . . 5
  - 1.5.1 Beweis durch beispiel . . . 5
  - 1.5.2 Wiederlegen von Behauptungen . . . 5
- 1.6 ∀ Aussagen beweisen . . . 6
  - 1.6.1 Ein Beispiel . . . 7
  - 1.6.2 Bemerkungen . . . 7

2 Verwenden von ∀ Aussagen **7**
- 2.1 Inferenzeregel um ∀ Aussagen zu verwenden . . . 7
  - 2.1.1 Beispiel . . . 7
  - 2.1.2 Beispiel . . . 8
- 2.2 Beispiel . . . 8

3 Einführung **9**
- 3.1 9

4 Fogen, Felder und Listen **10**
- 4.1 Verkettete Listen . . . 11
  - 4.1.1 Doppelt verkettete Listen . . . 11
  - 4.1.2 Einfach verkettete Listen . . . 14
- 4.2 Unbeschränkte Felder (Unbounded Arrays) . . . 14
  - 4.2.1 Amortisierte Komplexität unbeschr. Felder . . . 14
- 4.3 Amortisierte Analyse – allgemeiner . . . 14
- 4.4 Stapel und Schlangen . . . 15
- 4.5 Vergleich: Listen – Felder . . . 15

5 Hashing **16**
- 5.1 Hashing mit verketteten Listen . . . 16
- 5.2 Universelles Hashing . . . 16
- 5.3 Hashing mit Linearer Suche (Linear Probing) . . . 16

6 Sortieren **17**
- 6.1 Einfache Sortieralgorithmen . . . 18
- 6.2 Sortieren durch Mischen . . . 18
- 6.3 Untere Schranken . . . 18

7 Prioritätslisten **20**
- 7.1 20

8 Sortierte Listen **21**
- 8.1 21

9 Graphenreprarsentation **22**
9.1 22

10 Graphtravesierung **23**
10.1 23

11 Küreste Wege **24**
11.1 24

12 Minimale Spannbäume **25**
12.1 25

13 Optimierung **26**
13.1 26

1 Beweise

1.1 Statements

Ein **Statement/ Aussage** ist ein Mathematischer Ausdruck der entweder wahr oder Fallsch ist.

Beispiel:

- $2 \in \{x \in \mathbb{R} | x < 5\}$ (wahr)
- $3^2 + 5^2 = 8^2$ (falsch)

Dabei werden Ausdrüche wie $0 < x < 1$ verwendet um Mengen zu definieren.

$$A = \{x \in \mathbb{R} | 0 < x < 1\}$$

Wichtig ist hierbei der **Wahrheitswert** eines offenen Ausdrucks $0 < x < 1$ hängt von gewähltem x ab. Also ist

- $x = 1/2$ (wahr)
- $x = 5$ (falsch)

Die **Domäne** ist hierfür wichtig zu beachten. Für $\mathbb{N}$, gibt es kein x s.t. $0 < x < 1$, aber es gibt welche für $\mathbb{R}$

1.2 Formal mathematical proofs

Ein **formaler mathematischer Beweise** besteht aus einer nummerierten Sequent von **wahren Aussagen**. Jede Aussage in einem Beweis ist ein **Annahme** oder **folgt** aus vorherigen Aussagen dirch Ableitungsregel/ Inferenzregel (rule of inference). Die Letzte Aussage ist die die wir bewiesen haben.
$\Rightarrow$ Offene Aussagen können in Beweisen nicht auftreten!

Beispiel einer Inferenzregel: set definition rule

Wenn ein Element in einer Menge ist, dann können wir definierende Eienschaften ableiten. Andererseits, wenn es die definierende Eigenschaften erfüll, dann können wir ableiten das dasd Element in der Menge ist.

1.3 Set definition rule: Beispiel

Definiere $C = \{x \in \mathbb{R} | x < 2\}$
($x < 2 \wedge x \in \mathbb{R}$ ist die definierende Eigenschaft). Dabei gibt es zeiw Möglichkeiten für die Ableitung

Möglichkeit 1
1. $a \in C$
2. $a < 2 \wedge a \in \mathbb{R}(1; defC)$

Möglichkeit 2
1. $b < 2 \wedge b \in \mathbb{R}$
2. $b \in C(1; defC)$

Jede Aussage in dem Bewies hat eine Nummer. Wir begründen wie wir eine Aussage ableiten, zb $(1; defC)$ bedeutet wir leiten die aktuelle Aussage aus Aussage 1 mit der Definition von C und der set definition rule ab.

Bemerkung: $\wedge b \in R$

wird oft ausgelassen, wenn Kontext es zulässt.

1.4 Macro-steps in proofs

Problem:

Schauen wir uns folgenden Bewwis an:
(ass = Annahme (assumption) und prop = Eigenschaft(property))
Ist das ein akzrptabler Beweis? Akzeptanz von Makro-Schritten wie "**prop** $\mathbb{R}$"hängt von der Zielgruppe ab!
Welche Eigenschaft von $\mathbb{R}$ wurde benutzt?

1.5 Einfache Beweistechniken

1.5.1 Beweis durch beispiel

Beispiel: Zeigen Sie es gibt einePrimzahl zwischen 80 und 90.
Idee: Zeugen angeben für Primzahl (p) für die die Aussage gilt.

Beweis. Wähle $p = 83$ □

Ist das ausreichend?
Eigenflich, **NEIN**. Wie müssen noch **zeigen** das 83 tatsächlich eine Primzahl ist.
Das können wir tun in dem wir alle Teiler ausprobieren.

1.5.2 Wiederlegen von Behauptungen

Behauptung: Nimm an n ist eine Primzahl größer als 1. Dann ist $2^n - 1$ ebenfalls eine Primzahl.
Können Sie die Behauptung beweisen? Try hard $\cdots$
Wenn Sie es nicht können, dann sollen Sie darüber nachednken die Behauptung zu wiederlegen.

Eine Primzahl n für die $2^n - 1$ nicht prim ist, ist genug!
Das Gegenbeispiel ist $n = 11$ da 11 prim ist aber

$$2^{11} - 1 = 2047 = 23 \cdot 89$$

keine Primzahl ist!

1.6 $\forall$ Aussagen beweisen

Inferenzregel für definierte Beziehungen

The definition rule

Angenommen, es wurde eine Beziehung definiert. Wenn die Beziehung gilt (in irgendweinem Beweisschritt oder Annahme), dann kann die definierende Eigenschaft abgeleitet werden. Andererseits, wenn die definierende Eigenschaft gilt, dann kann die Beziehung abgeleitet werden.

Beispiel: Für Mengen A und B, definiere A ist **Teilmenge** von B, $A \subseteq B$, wenn **für alle** x **mit** $x \in A : x \in B$. Mit anderen Worten:
$A \subseteq B$ if and only if (iff) $\forall x((x \in A) \rightarrow (x \in B))$ ist wahr.

Möglichkeit 1:

1.	$A \subseteq B$	(ass 2)
2.	für alle x s.t.$x \in A : x \in B$	(1, def $\subseteq$)

Möglichkeit 2:

1.	für alle x s.t.$x \in A : x \in B$	(1, def $\subseteq$)
2.	$A \subseteq B$	(ass 2)

Inferenzeregel für $\forall$

Sei $\mathfrak{P}$ eine Formel. Beispielsweise steht $\mathfrak{P}(x)$ für $x \in A$ und $\mathfrak{Q}(x)$ steht für $x \in B$. Dann kann "**für alle x s.t.** $x \in A : x \in B$äls "**für alle x s.t.** $\mathfrak{P}(x) : \mathfrak{Q}(x)$"geschrieben werden.

Regeln um $\forall$ Aussagen zu beweisen (pr $\forall$)

Um Aussagen der Form "**für alle** x **in s.t.** $\mathfrak{P}(x) : \mathfrak{Q}(x)$", zu beweisen nimmt man an x sei **beliebig gewähtes Element (eigenvariable)** s.t $\mathfrak{P}(x)$ wahr ist. Dann zeige man $\mathfrak{Q}(x)$ ist wahr.

Generaliesirungen z.B. "für alle x, y s.t. $\mathfrak{P}(x, y) : \mathfrak{Q}(x, y)$"möglich

1.6.1 Ein Beispiel

Sei $C = \{x \in \mathbb{R} | x < 1\}$ und $D = \{x \in \mathbb{R} | x < 2\}$. Zeige $C \subseteq D$!
Wie können wir "für alle x s.t $\mathfrak{P}(x) : \mathfrak{Q}(x)$"wiederlegen?

1.6.2 Bemerkungen

- Durch Einrückung kennzeochnen wir **Teilbeweise** die von einer ANnahme wie SSei $x \in C$ beliebigäbhängen.
- Eine Annahme hat keine Begründung.
- Teilbeweise 2-4 basieren auf der Annahme in 1
- Schritte aus 1-4 können nicht in Begründungen auftauchen, sobald der Teilbeweis fertig ist (d. h, **nach pr $\forall$ in 5**)
- Wir schreiben oft "für alle $x \in C : x \in D$ßtatt "für alle $xs.t.x \in C : x \in D$"

2 Verwenden von $\forall$ Aussagen

2.1 Inferenzeregel um $\forall$ Aussagen zu verwenden

Die Regel um $\forall$ Aussagen in Beweisen zu verwenden (us $\forall$)

Wenn wir wissen das eine Aussage "für alle x s.t $\mathcal{P}(x) : \mathcal{Q}(x)$"wahr ist und wir $\mathcal{P}(t)$ berets als einen Schritt für eine Variable t im Beweis haben, dann können wir $\mathcal{Q}(t)$ ableiten.

2.1.1 Beispiel

1. $t \in A$
2. für alle x s.t.$x \in A : x \in B$
3. $t \in B$ $(1, 2; us\forall)$

2.1.2 Beispiel

1. $|a| < |b|$
2. für alle x s.t.$|x| \leq |y| : x^2 \leq y^2$
3. $a^2 < b^2$ $(1, 2; us\forall)$

2.2 Beispiel

Seiesn A, B, C Mengen. Zeige $\subseteq$ ist transitiv, d. h, zeige $A \subseteq B$ und $B \subseteq B$, dann $A \subseteq C$.

3 Einführung

3.1 ...

...

4 Fogen, Felder und Listen

Folgen spielen in der Informatik eine Überragende Rolle. Das sieht man schon an der Vielzahl von Begriffen: Folge, **Feld**, Schlange, **Liste**, Datei, Stapel, Zeichenkette, Log, $\cdots$. Wir unterscheiden:

- **abstrakter** Begriff [2, 3, 5, 7, 9, 11, $\cdots$] - Mathe
- Funktionalität (stack, $\cdots$) - Softwaretechnik
- Repräsentation - Algorithmik

Anwendungen $\cdots$

Form Follows Function

Operation	LIST	SLIST	UArray	CArray	explanation
$[\cdot]$	n	n	1	1	
$\lvert\cdot\rvert$	1^*	1^*	1	1	not with inter-list
FIRST	1	1	1	1	
LAST	1	1	1	1	
INSERT	1	1^*	n	n	InsertAfter only
REMOVE	1	1^*	n	n	RemoveAfter only
PUSHBACK	1	1	1^*	1^*	amortized
PUSHFRONT	1	1	n	1^*	amortized
POPBACK	1	n	1^*	1^*	amortized
POPFRONT	1	1	n	1^*	amortized
CONCAT	1	1	n	n	
SPLICE	1	1	n	n	
FINDNEXT	n	n	n^*	n^*	cache-efficient

4.1 Verkettete Listen

Eine verkettete Liste ist eine häufig verwendete Art von Datenstruktur, bei der jedes Element in der Liste Informationen über das nächste Element enthält. Es gibt verschiedene Arten von verketteten Listen, darunter einfach verkettete Listen und doppelt verkettete Listen.

4.1.1 Doppelt verkettete Listen

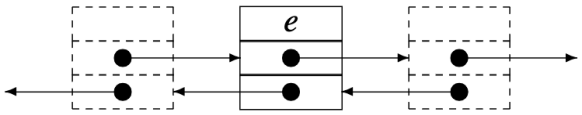
Bei doppelt verketteten Listen enthält jedes Listenelement Informationen sowohl über das vorherige als auch das nächste Element. Dies ermöglicht eine effiziente Navigation in beide Richtungen innerhalb der Liste.

In der Vorlesung haben wir eine Implementierung von doppelt verketteten Listen betrachtet. Jedes Listenelement wird durch die Klasse Ïtem"repräsentiert. Ein Item enthält zwei Handles (Zeiger) auf die vorherigen und nächsten Elemente in der Liste sowie ein Element selbst, das den eigentlichen Datenwert enthält.

Hier ist die Definition der Klasse Ïtem":

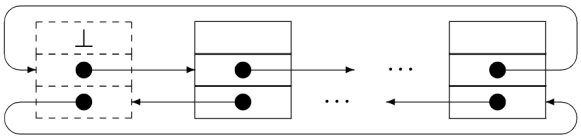
Listenglieder (Items)

```
Class Handle = Pointer to Item
Class Item of Element // one link in a doubly linked list
    e : Element
    next : Handle
    prev : Handle
    invariant next→prev = prev→next = this
```



Probleme:

- **Vogränger des ersten** Listenelements?
- **Nachfolger des letzten** Listenelements?



Trick: Dummy-Header Der Trick mit dem Dummy-Header ist ein nützliches Werkzeug zur Verbesserung der Implementierung und Verwendung von Listenstrukturen. Er steigert Lesbarkeit, Effizienz und Eleganz des Codes und erleichtert das Testen. Durch Vermeidung von Sonderfällen und Aufrechterhaltung der Invariante der verketteten Liste gewährleistet der Dummy-Header reibungslose und zuverlässige Listenoperationen.
Der Dummy-Header ist eine Technik zur Vereinfachung der Implementierung von Listenstrukturen. Durch Verwendung eines speziellen Listenelements am Anfang und Ende der Liste können viele Sonderfälle vermieden werden. Er dient als Platzhalter und Verweis für den Anfang und das Ende der Liste.
Der Dummy-Header bietet folgende Vorteile bei der Verwendung von Listen:

+ Die Invariante der verketteten Liste wird immer erfüllt. Der Dummy-Header stellt sicher, dass es immer einen Vorgänger für das erste Element und einen Nachfolger für das letzte Element gibt.

+ Durch Vermeidung vieler Sonderfälle wird die Implementierung einfacher, lesbarer und eleganter. Es müssen keine zusätzlichen Bedingungen für leere Listen oder Listen mit einem einzigen Element überprüft werden.

+ Der Zugriff auf das erste und letzte Element der Liste ist effizient und erfordert keine aufwändige Traversierung der gesamten Liste.

+ Der Dummy-Header erleichtert das Testen der Listenimplementierung, da die speziellen Fälle von leeren Listen und Listen mit einem einzigen Element bereits abgedeckt sind.

- Der zusätzliche Speicherplatz, der für den Dummy-Header benötigt wird. Dieser Faktor ist in der Regel vernachlässigbar, besonders bei längeren Listen.

Die Definition der Listenklasse mit dem Dummy-Header lautet:

```
class List of Element {
    // Item h ist der Vorgänger des ersten Elements und der Nachfolger des letzten Elem
    // Pos. vor jedem eigentlichen Element

    // !help

    // Einfache Zugriffsfunktionen
    Function head: Handle; return Adresse von h
    Function isEmpty: {0, 1}; return h.next = head // ⟨⟩?
    Function first: Handle; assert ¬isEmpty; return h.next
    Function last: Handle; assert ¬isEmpty; return h.prev
}
```

```
Procedure splice(a,b,t : Handle) {
    //!help
}
```

Der rest sind Einzeiler (?)

```
// Moving elements around within a sequence.
// ⟨... , a, b, c ... , a', c', ...⟩ ↦ ⟨... , a, c ... , a', b, c', ...⟩
Procedure moveAfter(b, a' : HANDLE) splice(b, b, a')
Procedure moveToFront(b : HANDLE) moveAfter(b, HEAD)
Procedure moveToBack(b : HANDLE) moveAfter(b, LAST)
```

Oder doch nicht? Speicherverwaltung! Die Speicherverwaltung in einer Programmiersprache kann potenziell sehr langsam sein. Ein Beispiel dafür ist eine Variable, die einmal erstellt wurde, wie zum Beispiel ein statisches Element in Java. Die *FREELIST* enthält ungenutzte Elemente, während *CHECKFREELIST* sicherstellt, dass sie nicht leer ist. In realen Implementierungen gibt es verschiedene Ansätze:

1. Ein naiver Ansatz, der jedoch eine gute Speicherverwaltung bietet.
2. Verfeinerte Konzepte zur Verwaltung der *FREELIST*, die über Klassen hinweg angewendet werden können und die Freigabe von Elementen ermöglichen.
3. Anwendungsspezifische Ansätze, zum Beispiel wenn man im Voraus weiß, wie viele Elemente insgesamt benötigt werden.

Items Löschen

```
// ⟨... , a, b, c, ...⟩ 7 ↦ ⟨... , a, c, ...⟩
Procedure remove(b : HANDLE) moveAfter( b, freeList.head)
Procedure popFront remove(FIRST)
Procedure popBack remove(LAST)
```

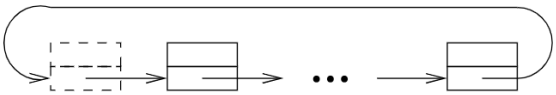
Elemente einfügen

Ganze (Teil)Listen Manipulieren

Suchen

Funktionalität ↔ Effizienz Wir betrachten folgendes Beispiel um Listenlängen zu verwalten. Neben den vorhandenen Elementen verfügt die Liste nun auch über eine zusätzliche Eigenschaft namens `SIZE`.
Problem: inter-list SPLICE geht nicht mehr in konstanter Zeit. Die Moral dabei ist das es keine perfekte Art, Listen zu implementieren gibt. Es hängt von den spezifischen Anforderungen und den gewünschten Operationen ab.

4.1.2 Einfach verkettete Listen



Vergleich mit doppelt verketteten Listen

- weniger Speicherplatz
- Platz ist oft auch Zeit
- eingeschränkter, z. B. kein remove
- merkwürdige Benutzerschnittstelle, z. B. removeAfter

Einfach verkettete Listen – Invariante?

Einfach verkettete Listen – SPLICE

Einfach verkettete Listen – PUSHBACK

Listen: Zusammenfassung, Verallgemeinerungen

Felder (Arrays)

Beschränkte Felder (Bounded Arrays)

4.2 Unbeschränkte Felder (Unbounded Arrays)

Unbeschränke Felder – Anwendungen

Unbeschränke Felder – Grundidee

Unbeschränke Felder mit teilweise ungenutztem Speicher

Kürzen

4.2.1 Amortisierte Komplexität unbeschr. Felder

Beweis: Konto-Methode (oder Versicherung)

4.3 Amortisierte Analyse – allgemeiner

Amortisierte Analyse – Diskussion

4.4 Stapel und Schlangen

4.5 Vergleich: Listen – Felder

Iterieren

Einfügen an zufälliger Position

Ausblick: Weitere Repräsentationen von Folgen

5 Hashing

...

5.1 Hashing mit verketteten Listen

Analyse

Zufällige

5.2 Universelles Hashing

idee: Nutze nur betrimmre

Eine einfache universelle Familie

m ei eine primzahl, $Key \subseteq \{0, \cdots, m-1\}^k$
Satz 3 Für $a = (a_1,)$

5.3 Hashing mit Linearer Suche (Linear Probing)

Offenees Hashing

Der einfache Teil

```
your
code
example
```

Remove

```
your
code
example
```

Verketten $\leftrightarrow$ Lineare Suche

Volllaufen: Verketten weniger empfindlich. Unbeschränktes **offenes** Hashing hat nur amortisiert konstante Einfügezeit.

Cache: Lineare Suche besser. Vor allem **DOALL**

Zeit Abwägung: Kompliziert! Abhämgig von n, **Füllgrad**, **Elementgröße**,

6 Sortieren

Formaler

Gegeben: Elementfolge $s = [e_1, \cdots, e_n]$
Gesucht: $s' = [e'_1, \cdots, e'_n]$

- s' ist Permutation von s
- $e'_1 \leq \cdots \leq e'_n$ für eine **lineare Ordnung** $' \leq'$

Anwendungsbeispiele

- Allgemein: Vorverarbeitung
- Suche: **Telefonbuch** $\leftrightarrow$ unsortierte Liste
- Gruppieren (Alternative Hashing?)

Beispiel aus Kurs/ Buch

- Aufabau von Suchbäumen
- Kruskals MST- Algorithmus
- Verarbeitung von Intervallgraphen (z.B Hotelbuchungen)
- Rucksackproblem
- Scheduling, die schwersten Probleme zuerst
- Sekundärspeicheralgorithmen, z.B Datenbank-Join

Viele verwandte Probleme. Zum Beisoiel **Transposition** dünner Matrizen, **invertierten Index** aufbauen, Konversion zwischen Graphrepräsentationen.

Überblick

- Einfache Algorithmen/ kleine Datenmengen
- **Mergesort** - ein erster effizienter Algorithmus
- Eine passende **untere Schranke**
- **Quicksort**
- das Asuwahlproblem
- ganzzahlige Slüssel - jenseits der unteren Schranke

6.1 Einfache Sortieralgorithmen

Sortieren durch Einfügen (inster sort)

```
procedure MyAlgorithm(x)
    y ← 0
    for i ← 1 to 10 do
        if i is odd then
            y ← y + x
        end if
    end for
    return y
end procedure
```

Sentinels am Beispiel Sortieren durch Einfügen

```
procedure insertSort(a : Array [1..n] of Elements)
    for i ← 2 to n do do
        invariant a[1] ≤ ... ≤ a[i − 1] ···
    end for
    return y
end procedure
```

Analyse

6.2 Sortieren durch Mischen

idee: Teile und Hersche

```
procedure mergSort([e_1, ··· , e_n] : Sequence of Elements)
    if n = 1 then
        return [e_1]
    end if
end procedure
```

Mischen(merge)

Beispiel

Mischen

Analyse

$$T(n) = O(n) + T(\lceil n/w \rceil) + T(\lfloor n/s \rfloor)$$

Problem: Runderei

6.3 Untere Schranken

geht es schneller als $\Theta(nlogn)$?

Eine vergleisbasiere untere Schranke

Vergkeichsbasiertes Sortieren: Informationen über Elemente nur durch Zwei-Wege-Vergleich $e_i \leq e_j$?

Satz: Deterministische vergleichsbasierte Sortieralgorithmen brauchen

$$\mathbf{n \log n} - O(n)$$

Vergleiche im schlechtesten Fall.

Beweis: Betrachte Eingaben, die Permutatuiónen von $1..n$ sin. Es gibt genau $n!$ solche Permutationen.

Baumbasierte Sortierer-Darstellung

Beweis

Baum der **Tiefe T** hat höchstens 2^T Blätter.

$$\Rightarrow 2^T \geq n!$$

$$\Leftrightarrow T \geq \log n! \geq \log(\frac{n}{e})^n = n \log n - n \log e = n \log n - O(n)$$

einfache Aproximation der Fakultät: $(\frac{n}{e})^n \leq n! \leq n^n$
Beweis für den **linken Teil:**

$$\ln n! = \sum_{2 \leq i \leq n} \ln i \geq \int_1^n \ln x \, dx = [x(\ln x - 1)]_{x=1}^{x=n} \geq n(\ln n - 1)$$

$$n! \geq e^{n(\ln n - 1)} = \frac{e^{n \ln n}}{e^n} = \frac{n^n}{e^n} = (\frac{n}{e})^n$$

Randomisierung, Mittlere Auführungszeit

Satz: immer noch $n \log n - O(n)$ Vergleiche.
Beweis: nicht hier.

Quicksort - erster Versuch

Idee: Teile-und-Hersche aber vergleichen mit mergsort ändersrum Leiste Arbeit **vor** rekursivem Aufruf

Quicksort - Analyse im schlechtesten Fall

Annahme: Pivot ist immer Minimum (oder Maximum) der Eingaben

$$T(n) = \begin{cases} \Theta(1) & \text{if } n = 1 \\ \Theta(n) & \text{if } n \geq 2 \end{cases}$$

$\Rightarrow T(n) = \Theta(n + (n-1) + \cdots + 1) = \Theta(n^2)$

Schlechteser Fall: Beispiel

Quicksort - Analyse im besten Fall

7 Prioritätslisten

7.1 ...

...

8 Sortierte Listen

8.1 ...

...

9 Graphenreprarsentation

9.1 ...

...

10 Graphtravesierung

10.1 ...

...

11 Küreste Wege

11.1 ...

...

12 Minimale Spannbäume

12.1 ...

...

13 Optimierung

13.1 ...

...