Probeklausur

Datenstrukturen und Algorithmen

08.07.2013, Prof. Dr. L. Kobbelt

Nachname: _____

Vorname: _____

Matrikelnummer: _____

Studiengang: _____

- Hinweis: Diese Klausur entspricht einer Auswahlklausur, bei welcher 75% der erreichbaren Punkte für die Note 1.0 ausreichen. Die diesjährige Klausur wird keine Auswahlklausur sein.
- Schreiben Sie auf jedes Blatt **Vorname**, **Name** und **Matrikelnummer**.
- Geben Sie Ihre Antworten in lesbarer und verständlicher Form an.
- Bitte beantworten Sie die Aufgaben auf den **Aufgabenblättern** (benutzen Sie auch die Rückseiten). Antworten auf anderen Blättern können nur berücksichtigt werden, wenn **Name, Matrikelnummer und Aufgabennummer** deutlich darauf erkennbar sind.
- Was nicht bewertet werden soll, kennzeichnen Sie bitte durch **Durchstreichen**.
- Werden Täuschungsversuche beobachtet, so wird die Klausur mit **nicht bestanden** bewertet.
- Geben Sie am Ende der Bearbeitungszeit **alle Blätter zusammen mit den Aufgabenblättern ab**.
- Verwenden Sie ausschließlich dokumentenechte Stifte mit schwarzer oder blauer Farbe. Insbesondere sind keine Bleistifte und keine Stifte mit roter Farbe zugelassen.

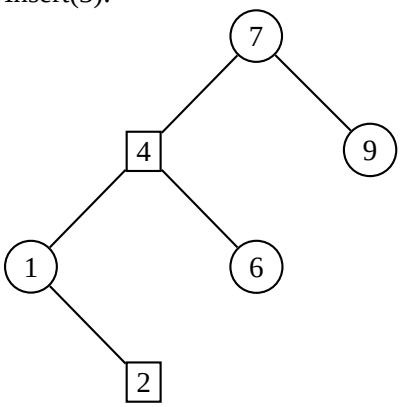
	Thema	Punkte	Erreichte Punkte
1	Bäume	14	
2	ADT	11	
3	Graphen	11	
4	Sortieren	8	
5	Multiple Choice	15	
6	Dynamisches Programmieren	8	
7	Hashing	5	
8	Laufzeitanalyse	8	
	Summe	80	

Vorname	Name	Matr.-Nr.	Seite
			1 / 27

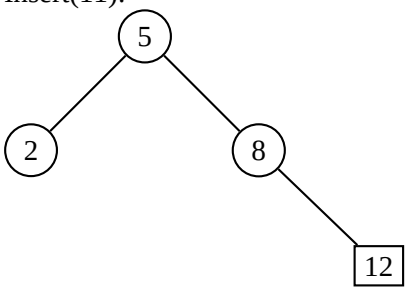
Aufgabe 1: Bäume

1. (3 Punkte) Fügen Sie in die folgenden Rot-Schwarz-Bäume jeweils die angegebenen Schlüssel ein. Zeichnen Sie schwarze Knoten als Kreise und rote Knoten als Quadrate.

   Insert(3):

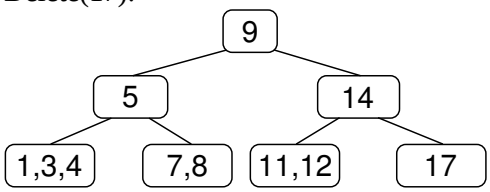


   Insert(11):

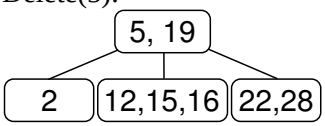


2. (4 Punkte) Löschen Sie aus den folgenden B-Bäumen der Ordnung $t = 2$ jeweils die angegebenen Schlüssel.

   Delete(17):



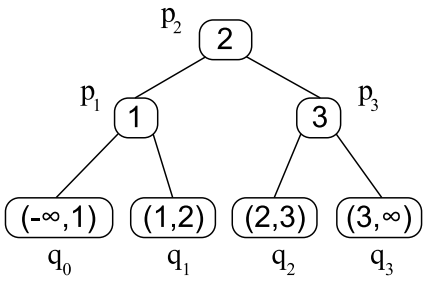
   Delete(5):



3. (2 Punkte) Beschreiben Sie den “One-Pass-Algorithmus” zum Einfügen von Knoten in einen B-Baum der Ordnung t. Welches Problem des naiven Einfüge-Algorithmus verhindert der One-Pass-Algorithmus?

4. (2 Punkte) Geben Sie alle korrekten B-Bäume der Ordnung $t = 2$ an, die die Schlüssel $2, 4, 6, 8, 10$ enthalten. Hierbei sollen auch Bäume berücksichtigt werden, die durch das Einfügen und spätere Löschen zusätzlicher Elemente entstehen können.

5. Wir betrachten den folgenden Suchbaum:



Gegeben seien die Wahrscheinlichkeiten der inneren Knoten $p_1 = 0.25$, $p_2 = 0.20$, $p_3 = 0.15$ und der Blätter $q_0 = 0.25$ sowie $q_1 = q_2 = q_3 = 0.05$.

(a) (1 Punkt) Berechnen Sie die erwarteten Kosten (erwartete Anzahl von Knotenzugriffen) für den oben angegebenen Suchbaum.

(b) (2 Punkt) Geben Sie den optimalen Suchbaum für die angegebenen Wahrscheinlichkeiten an. Wieviel Kosteneinsparung kann erwartet werden?
(Hinweis: Zusammen mit dem Baum aus Teil (a) existieren 5 mögliche Bäume. Gesucht ist der Baum mit den optimalen Kosten.)

Aufgabe 2: Abstrakte Datentypen

In dieser Aufgabe soll ein abstrakter Datentyp PLISTE spezifiziert werden. Bei diesem Datentyp handelt es sich um eine gepackte Liste, d.h. eine Liste, deren Elemente Paare von Werten und Anzahlen sind. So wird z.B. die normale Liste mit den vier Buchstaben $[A, B, B, A]$ durch die gepackte Liste $[(A, 1), (B, 2), (A, 1)]$ dargestellt. Eine solche Kodierung nennt man "Run-Length Encoding" (kurz: RLE).

1. (2 Punkte) Zuerst benötigen wir einen Hilfstypen LISTE für normale (ungepackte) Listen von Werten.

   Definieren Sie den ADT LISTE für eine LIFO-Liste mit den folgenden Operationen durch ihre Signatur und geeignete Axiome:

$Create$	Erzeugt eine leere Liste.
$Push$	Fügt ein Element am Anfang der Liste ein.
Top	Liefert das erste Element der Liste zurück.
Pop	Entfernt das erste Element der Liste.
$IsEmpty$	Liefert $True$, falls die Liste leer ist.

   Sie können davon ausgehen, dass der Elementtyp WERT bereits definiert ist. Ebenso dürfen Sie davon ausgehen, dass der Typ BOOLEAN für Wahrheitswerte mit $True, False\colon \rightarrow$ BOOLEAN gegeben ist.

2. (2 Punkte) Nun definieren wir den ADT PLISTE für gepackte LIFO-Listen.

   Die folgenden Operationen sollen auf einer PLISTE durchführbar sein:

$PCreate()$	Erzeugt eine leere gepackte Liste.
$PPush()$	Fügt einen Wert und die zugehörige Anzahl an den Anfang der gepackten Liste ein.
$Value$	Liefert den im ersten Element der Liste gespeicherten Wert zurück.
$Count$	Liefert die im ersten Element der Liste gespeicherte Anzahl zurück.
$PPop$	Entfernt das erste Element (also das erste Paar) der gepackten Liste.
$PIsEmpty$	Liefert $True$, falls die gepackte Liste leer ist.

   Sie können davon ausgehen, dass der Elementtyp WERT bereits definiert ist. Außerdem dürfen Sie wieder BOOLEAN benutzen und können davon ausgehen, dass der Typ NAT für natürliche Zahlen wie folgt gegeben ist:

$$\begin{array}{rlcl} Zero: & & \rightarrow & \text{NAT} \\ Succ, Pred: & \text{NAT} & \rightarrow & \text{NAT} \\ IsZero: & \text{NAT} & \rightarrow & \text{BOOLEAN} \end{array}$$

$$\begin{aligned} Pred(Succ(n)) &= n \\ IsZero(Zero()) &= True() \\ IsZero(Succ(n)) &= False() \end{aligned}$$

   Geben Sie die Signatur von PLISTE sowie geeignete Axiome an.

3. Es soll eine Operation $UnRLE$ definiert werden, die eine gepackte Liste in eine normale (ungepackte) Liste überführt. Dabei soll der Wert jedes Elements entsprechend der angegebenen Anzahl vervielfältigt werden. So soll z.B.

$$UnRLE(PPush(A, 1, PPush(B, 2, PPush(A, 1, PCreate()))))$$

zu

$$Push(A, Push(B, Push(B, Push(A, Create()))))$$

ausgewertet werden können.

(a) (2 Punkte) Geben Sie Signatur und Axiome für $UnRLE$ sowie evtl. benötige Hilfsoperationen an. Die Datentypen WERT, LISTE, BOOLEAN und NAT sowie ihre Axiome dürfen vorausgesetzt werden.

(b) (2 Punkte) Geben Sie Pseudocode für die Operation $UnRLE$ an. Benutzen Sie hierfür höchstens die Operationen $Create$, $Push$, $PPush$, $PPop$, $Value$, $Count$, $PIsEmpty$, $IsZero$ und $Pred$.

4. (3 Punkte) Es soll eine Operation RLE definiert werden, die eine normale (ungepackte) Liste bekommt und eine gepackte Liste liefert, so dass die Originalliste durch entsprechendes Vervielfältigen wieder hergestellt werden kann. So soll z.B.

$$RLE(Push(A, Push(B, Push(B, Push(A, Create())))))$$

zu

$$PPush(A, 1, PPush(B, 2, PPush(A, 1, PCreate())))$$

ausgewertet werden können.

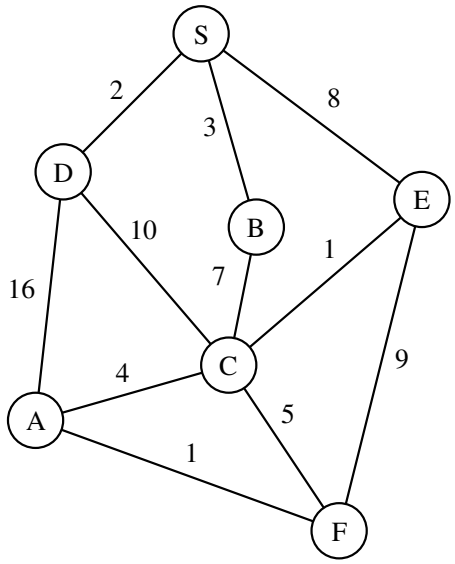
Die Signaturen von RLE und einer benötigten Hilfsfunktion $RLE2$ sind

$$\begin{array}{rrcl} RLE: & \text{LISTE} & \rightarrow & \text{PLISTE} \\ RLE2: & \text{LISTE} \times \text{WERT} \times \text{NAT} & \rightarrow & \text{PLISTE} \end{array}$$

Geben Sie Axiome für RLE und $RLE2$ an. Die Datentypen WERT, LISTE, BOOLEAN und NAT sowie ihre Axiome dürfen vorausgesetzt werden.

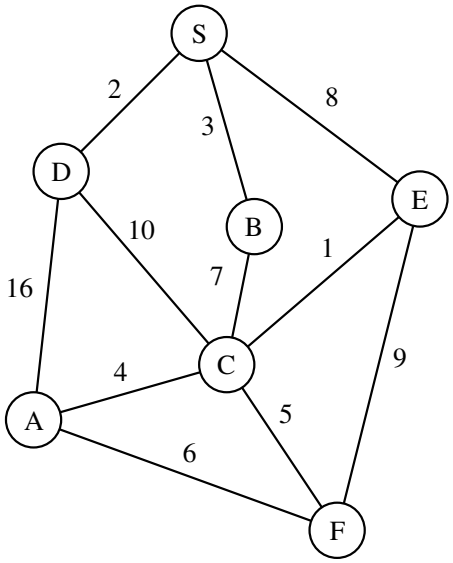
Aufgabe 3: Graphen

1. (2 Punkte) Wenden Sie Dijkstras Algorithmus auf den folgenden Graphen G mit Startknoten S an. Geben Sie die Werte aller oberen Schranken nach jedem Durchlauf der Hauptschleife von Dijkstras Algorithmus in Form einer Tabelle an. Die Tabelle enthält also eine Spalte für jeden Knoten $v \in G$. Pro Iteration soll eine Zeile hinzugefügt werden, die die aktuellen oberen Schranken $\delta(S, v)$ enthält. Markieren Sie dabei in jeder Zeile den Knoten, den Sie auswählen.



	S	A	B	C	D	E	F	Vertex

2. (1 Punkt) Wenden Sie den Algorithmus von Prim an, um einen minimal spannenden Baum des folgenden Graphen zu berechnen. Geben Sie dazu die Kanten in der Reihenfolge an, in der sie zur Lösungsmenge hinzugefügt werden. Da alle Kantengewichte verschieden sind, genügt es, jede Kante durch ihr Gewicht zu identifizieren.



3. Union-Find Strukturen

   (a) (1 Punkt) Führen Sie die folgenden Operationen mit Höhenbalancierung auf einer anfangs leeren Union-Find Struktur aus. Beachten Sie: Wenn die Höhe der Bäume gleich ist, wird der Repräsentant mit dem kleineren Wert der neue Repräsentant. Zeichnen Sie den entstandenen Wald nach den Operationen

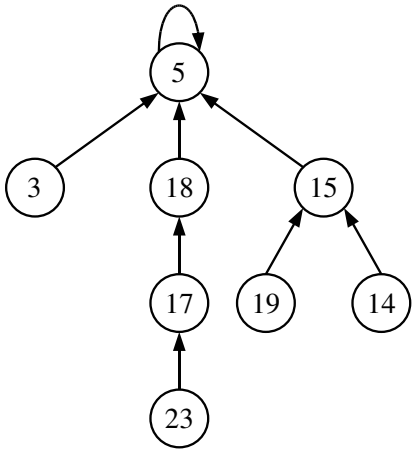
   - ```
     for i = 1 to 5 do
         MakeSet(i)
     ```

   - ```
     Union(1, 3)
     Union(2, 5)
     ```

   - ```
     Union(4, 5)
     ```

   - ```
     Union(5, 3)
     ```

(b) (1 Punkt) Führen Sie auf der Union-Find Struktur



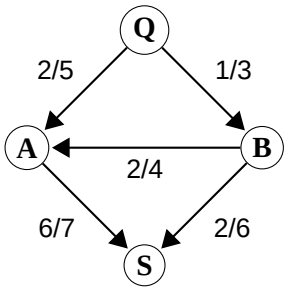
nacheinander die folgenden Operationen mit Pfadkompression aus. Zeichnen Sie nach jeder Operation den entstandenen Baum.

- `Find(14)`

- `Find(17)`

(c) (2 Punkte) Geben Sie den Kruskal-Algorithmus in Pseudo-Code mit Hilfe von Union-Find Strukturen an. Verwenden Sie dazu die Funktionen `Emptyset`, `MakeSet`, `Find` und `Union`.

4. (4 Punkte) Wenden Sie den Algorithmus von *Dinic* auf den folgenden Graphen G an. Geben Sie zu jeder Iteration jeweils den Fluss in G, den Restgraph und das Niveaunetzwerk an.



Aufgabe 4: Sortieren

1. (2 Punkte) Sortieren Sie die folgende Zahlenfolge aufsteigend mit Hilfe des Algorithmus Merge-Sort aus der Vorlesung. Verwenden Sie hierbei die iterative Variante. Geben Sie den Inhalt des Arrays nach dem Zusammenführen von allen Teilfolgen der gleichen Länge an.

13	56	77	42	14	21	83	64	22	98	59	11	85	51	19	73

2. (3 Punkte) Für die vergleichsbasierte Sortierung von n Werten wurde in der Vorlesung gezeigt, dass man mindestens Zeitaufwand $\Omega(n \log n)$ benötigt. Wieso kann ein Algorithmus wie Radix-Sort dann in $O(n)$ laufen?

3. (3 Punkte) Gegeben seien n Punkte $(x_i, y_i) \in \mathbb{R}^2$ im Quadrat mit den Ecken $(1, 0)$, $(0, 1)$, $(-1, 0)$ und $(0, -1)$, d.h., $|x_i| + |y_i| \leq 1$. Diese Punkte seien gleichverteilt, d.h. die Wahrscheinlichkeit, dass ein Punkt in einem Gebiet G innerhalb des Quadrates zu liegen kommt, ist proportional zum Flächeninhalt von G.

   Beschreiben Sie eine Variante des Bucketsort-Algorithmus, der die Punkte nach ihrer Betragssumme $|X| + |Y|$ sortiert und dessen Laufzeit im erwarteten Fall $O(n)$, d.h. linear ist. Geben Sie eine Definition für die verwendeten Buckets sowie eine Begründung für die lineare Laufzeit an.

Aufgabe 5: Multiple Choice

Bei einigen der folgenden Fragen sind mehrere der vorgegebenen Antworten richtig. Kreuzen Sie alle richtigen Antworten an. Pro Frage erhalten Sie für eine vollständig richtige Antwort 1 Punkt. Beantworten Sie ein Frage nicht, so erhalten Sie 0 Punkte. Beantworten Sie eine Frage unvollständig oder falsch, so erhalten Sie -0.5 Punkte. Insgesamt können Sie jedoch in dieser Aufgabe keine negative Punktzahl erhalten.

1. Gegeben sei folgende Rekursionsgleichung

$$T(1) = 7$$
$$T(n) = 7 \cdot T(\frac{n}{7}) + 7 \cdot n \quad \text{für } n > 1$$

Wie lässt sich T dann klassifizieren?

◯ $T = \Omega(1)$
◯ $T = O(n \log n)$
◯ $T = O(n^3)$
◯ Keine der obigen Klassifizierungen trifft zu.

2. Gegeben sei folgende Rekursionsgleichung

$$T(1) = 42$$
$$T(n) = 4 \cdot T(\frac{n}{2}) + 8 \cdot n^2 + 16 \cdot n \quad \text{für } n > 1$$

Wie lässt sich T dann klassifizieren?

◯ $T = O(n \log n)$
◯ $T = O(n^2 \log n)$
◯ $T = \Omega(n^2)$
◯ Keine der obigen Klassifizierungen trifft zu.

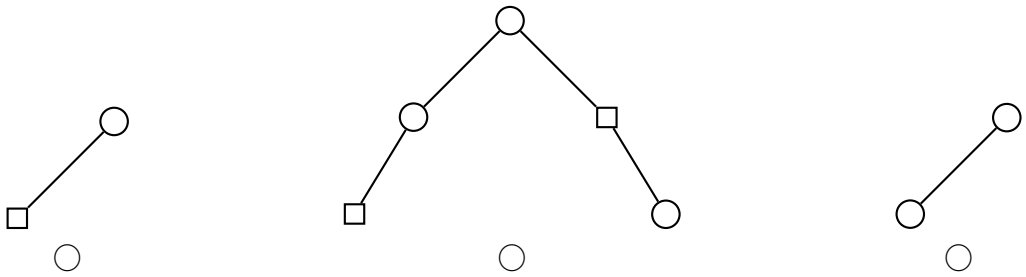
3. In zusammenhängenden Graphen $G = (V, E)$ gilt $|E| = \Omega(|V|)$.

◯ Richtig ◯ Falsch

4. Zwei verschiedene spannende Bäume eines Graphen enthalten mindestens eine gemeinsame Kante.

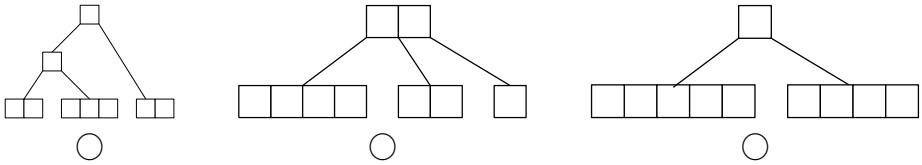
◯ Richtig ◯ Falsch

5. Welche der folgenden Bäume sind Rot-Schwarz-Bäume? Kreise stellen schwarze Knoten dar, Quadrate stellen rote Knoten dar.



◯ ◯ ◯

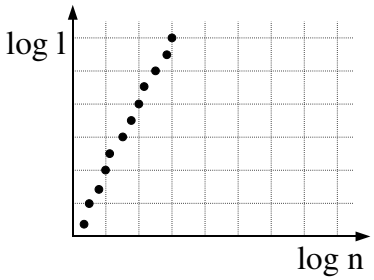
6. Welche der folgenden Bäume können B-Bäume sein?



7. Es gibt einen planaren, ungerichteten Graphen mit 5 Knoten und 10 (verschiedenen) Kanten.

◯ Richtig ◯ Falsch

8. Die Laufzeiten l_i eines Algorithmus A wurden für eine Menge von Datensätzen der Größen n_i gemessen. Es ergab sich folgender Zusammenhang, wobei sowohl die n- als auch die l-Achse eine logarithmische Skala haben.



Welche asymptotische Komplexität hat A vermutlich?

◯ $\Theta(n)$ ◯ $\Theta(n \log n)$ ◯ $\Theta(n^2)$

9. Bei geschlossenem Hashing beträgt der erwartete Aufwand zum Einfügen eines Elements in eine Hashtabelle mit Belegungsfaktor α

◯ $O(\alpha^2)$ ◯ $O(1+\alpha)$ ◯ $O(\frac{1}{1-\alpha})$

10. Die minimale Zahl an Kopieroperationen, um eine Folge der Länge n mit Selection-Sort zu sortieren, ist

◯ $\Theta(n)$ ◯ $\Theta(n \log n)$ ◯ $\Theta(n^2)$

11. Jeder ungerichtete Graph $G = (V, E)$ mit mehr als $|V|$ Kanten ist zusammenhängend.

◯ Richtig ◯ Falsch

12. Das *Versickern* eines Elements bei Heap-Sort benötigt $\Omega(n \log n)$ Zeitaufwand.

◯ Richtig ◯ Falsch

13. Dijkstras Algorithmus hat auf einem Baum mit n Knoten einen Aufwand von

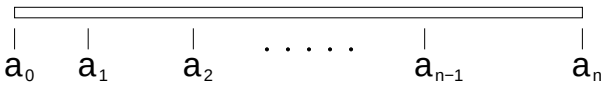
◯ $\Theta(n)$ ◯ $\Theta(n \log n)$ ◯ $\Theta(n^2)$

14. Die Komplexität des Algorithmus von Dinic auf einem Graphen $G = (V, E)$ ist

◯ $O(|V|^2 \cdot |E|)$ ◯ $O(|V|^3)$ ◯ $O(|V|^4)$

Aufgabe 6: Dynamisches Programmieren

Gegeben sei ein Stab der Länge $l \in \mathbb{N}_{>0}$. Dieser Stab soll an $n-1$ Stellen $a_i \in \mathbb{N}, 0 \leq a_i \leq l, i = 1, \ldots, n-1$ in n Segmente zerteilt werden. Der Einfachheit halber setzen wir außerdem $a_0 = 0$ und $a_n = l$ und es gelte $a_i < a_{i+1}\ \forall i$.



Die Kosten für einen Schnitt durch ein Segment der Länge m betragen m. Die Kosten, um den Stab an den vorgegebenen Stellen zu zerschneiden, hängen von der Reihenfolge der Schnitte ab. Beispiel: Ein Stab der Länge 10 soll an den Stellen $a_1 = 5$, $a_2 = 8$ und $a_3 = 9$ zerschnitten werden. Schneidet man in der Reihenfolge a_1, a_2, a_3 so betragen die Kosten $10 + 5 + 2 = 17$. Schneidet man in der Reihenfolge a_3, a_2, a_1 so betragen die Kosten hingegen $10 + 9 + 8 = 27$.
In dieser Aufgabe soll ein auf Dynamischem Programmieren basierender Algorithmus hergeleitet werden, der die minimalen Kosten berechnet, die anfallen, wenn der Stab zerschnitten werden soll. Hierzu bezeichne $c(i, j), 0 \leq i < j \leq n$ die minimalen Kosten um alle Schnitte zwischen a_i und a_j durchzuführen. Gesucht ist also $c(0, n)$.

1. Geben Sie eine Rekursionsformel für $c(i, j)$ an.

   Basisfall: (1 Punkt) Für alle i mit $0 \leq i < n$ ist

   $$c(i, i+1) =$$

   Rekursion: (3 Punkte) Für alle i, j mit $0 \leq i$ und $i + 1 < j$ und $j \leq n$ ist

   $$c(i, j) =$$

2. (2 Punkte) Geben Sie in Pseudocode einen auf Dynamischem Programmieren basierenden Algorithmus an, der $c(0, n)$ berechnet.

3. (2 Punkte) Geben Sie eine kleinste obere Schranke für die asymptotische Laufzeit Ihres Algorithmus an (mit kurzer Begründung).

Aufgabe 7: Hashing

1. Erstellen Sie durch sukzessives Einfügen von Datensätzen mit den Schlüsselwerten

$$7, 19, 24, 63, 50, 73, 37, 54, 43, 49, 44$$

eine Hashtabelle der Länge $m = 11$. Verwenden Sie dazu die Hashfunktion

$$h(x) = x \mod 11$$

und

(a) (1 Punkt) geschlossenes Hashing mit linearem Sondieren

0	1	2	3	4	5	6	7	8	9	10

(b) (1 Punkt) geschlossenes Hashing mit quadratischem Sondieren. Verwenden Sie die Variante aus der Vorlesung, bei der die Quadrate addiert werden.

0	1	2	3	4	5	6	7	8	9	10

2. Betrachten Sie nun offenes Hashing, bei dem Kollisionen durch Verkettung von Elementen behandelt werden.

   (a) (1 Punkt) Geben Sie die worst-case Laufzeiten (in O-Notation) für die erfolgreiche Suche, die erfolglose Suche, das Einfügen und Löschen in einer offenen Hashtabelle mit regulären (unsortierten) Listen an.

   (b) (2 Punkte) Die Listen seien nun sortiert. Geben Sie wieder die worst-case Laufzeiten für die Operationen erfolgreiche Suche, erfolglose Suche, Einfügen und Löschen an. Welche Variante der Listen ist effizienter im worst-case?

Aufgabe 8: Laufzeitanalyse

1. (2 Punkte) Zeigen oder widerlegen Sie:
   $f(n) \in O\big(h(n)\big) \wedge g(n) \in O\big(h(n)\big) \Rightarrow f(n) \cdot g(n) \in O\big(h(n) \cdot h(n)\big)$

2. (2 Punkte) Zeigen oder widerlegen Sie: $a^{2n} \in \Theta(a^n)$ für eine Konstante $a > 1$.

3. (2 Punkte) Zeigen oder widerlegen Sie: $\ln(n^2) \in O(\ln(n))$.

4. (2 Punkte) Es seien $f(n) := \ln(\ln(n))$ und $g(n) := (\ln(n))^2$.
   Gilt $f \in O(g)$? Gilt $g \in O(f)$? Beweisen Sie Ihre Behauptung.

Zusatzblatt 1

Vorname	Name	Matr.-Nr.	Seite
			27 / 27

Zusatzblatt 2