



Modul
Grundlagen von Datenbanken
Wintersemester 2020/2021

Probeklausur
??.??..2021, ?? Uhr, digital

--

Aufgabe	1	2	3	4	5	6	7	8	9	10	Gesamt
Punkte	18	12	12	14	12	6	8	12	11	15	120
Ergebnis											

Rechtsmittelbelehrung:

Gegen die Bewertung dieser Prüfungsleistung kann innerhalb eines Monats nach ihrer Bekanntgabe Widerspruch erhoben werden. In diesem Zeitraum kann die Bewertung der Klausur eingesehen werden. Der Widerspruch ist schriftlich oder zur Niederschrift beim Vorsitzenden des B.Sc./M.Sc.-Prüfungsausschusses einzulegen. Es wird darauf hingewiesen, dass ein erfolgloses Widerspruchsverfahren kostenpflichtig ist.

Bearbeitungshinweise:

- Die Bearbeitungsdauer der Klausur beträgt insgesamt 120 Minuten.
- Es werden maximal 120 Punkte vergeben. Dabei entsprechen die den einzelnen Aufgaben zugeordneten Punkte den jeweils vorgesehenen Bearbeitungszeiten (in Minuten). Zum Bestehen der Klausur sind 50% der erreichbaren Punkte hinreichend.
- Es sind alle Hilfsmittel zugelassen, die keine Interaktion mit anderen Personen erfordern.

Viel Erfolg!

Matrikel-Nr:

Name:

Hiermit versichere ich, dass ich die Klausur selbstständig und unter ausschließlicher Nutzung der erlaubten Hilfsmittel bearbeitet und die Rechtsmittelbelehrung zur Kenntnis genommen habe.

Unterschrift: _____

Blättern Sie erst weiter, wenn die Bearbeitungszeit offiziell begonnen hat.

Aufgabe 1: Informationsmodellierung mit dem Entity-Relationship-Modell

[18 P.]

Erfassen Sie die im Folgenden beschriebenen Informationsstrukturen (Ausschnitt einer Anwendung zur Verwaltung der Patienten und deren Behandlungen in einem Ärztehaus) in einem ER-Diagramm. Beziehen Sie sich dabei genau auf die gegebene Beschreibung, ohne weiteres Wissen zu möglicherweise ähnlichen Anwendungsbereichen einfließen zu lassen. Markieren Sie in Ihrem Entwurf Primärschlüssel durch Unterstreichung und konkretisieren Sie die Abbildungstypen durch Kardinalitätsrestriktionen (Notation: [min,max]). Verwenden Sie unbedingt die aus der Vorlesung bekannte Notation. Andere Notationen werden nicht gewertet! Ausnahme: Die Zuordnung bei 1:n Abbildungstypen bleibt Ihnen überlassen, muss aber eindeutig als solche markiert werden (z.B. durch ein ausformuliertes Beispiel). Benutzen Sie möglichst wenige Entitäten (Ausnahme Vererbung). Etwaige (Existenz-)Abhängigkeiten sollen nicht modelliert werden.

Es werden Angestellte und Patienten sowie deren Behandlungen betrachtet. Die Angestellten untergliedern sich in Hilfskräfte, welche eine Ausbildung haben, und Ärzte, welche ein Fachgebiet haben. Beiden Angestellten-Gruppen gemein ist, dass für sie ein Gehalt, ein Einstellungsdatum sowie ein Vertragsende festgelegt ist. Angestellte und Patienten haben personenspezifische Eigenschaften: eine eindeutige Identifikationsnummer (PID), einen Familiennamen, einen Vornamen und ein Geburtsdatum. Jede Hilfskraft kann (maximal) einem Arzt zugeordnet sein. Einem Arzt können wiederum mehrere Hilfskräfte zugeordnet werden. Für Patienten werden neben den bereits genannten personenspezifischen Eigenschaften noch die Größe, das Gewicht, die Adresse, welche sich aus einer Straße, einer Hausnummer, einer PLZ und einem Ort zusammensetzt, und der Name der Krankenkasse gespeichert.

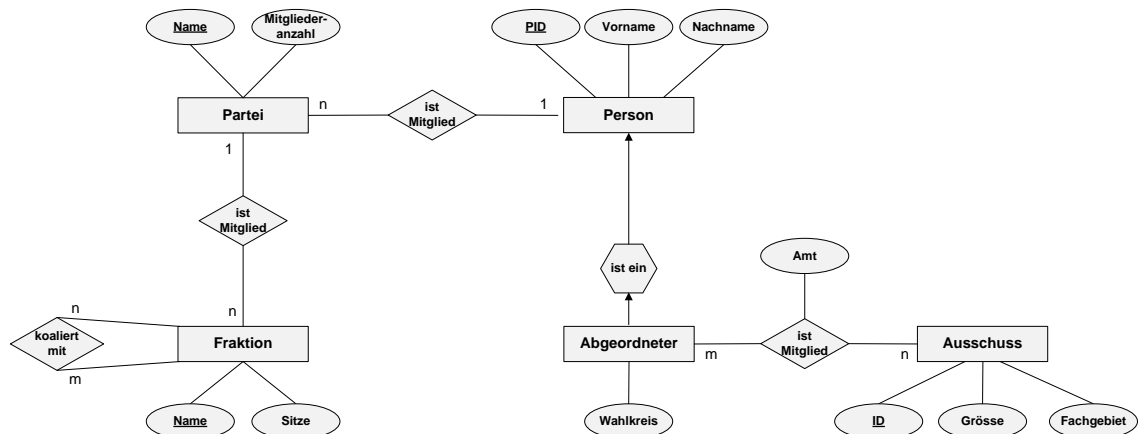
Eine Krankheit hat einen eindeutigen Identifikationscode (WHOID), einen Namen und eine Reihe von Symptomen. Ein Medikament hat eine eindeutige Zulassungsnummer (ZNr) und einen Namen. Ein Medikament kann gegen eine oder mehrere Krankheiten einsetzbar sein. Zwar kann es für jede Krankheit mehrere geeignete Medikamente geben, es ist aber auch möglich, dass eine Krankheit nicht durch Medikamente behandelbar ist. Zu beachten ist, dass ein Medikament Wechselwirkungen mit anderen Medikamenten haben kann, aber nicht haben muss.

Eine Behandlung ist durch eine eindeutige Behandlungsnummer (BehNr) gekennzeichnet. Zudem besitzt sie ein Datum an dem sie stattfand. Eine Behandlung betrifft immer genau einen Patienten, einen Arzt und eine Krankheit. Ein Patient kann mehrmals (aber mindestens einmal) behandelt werden, ein Arzt kann mehrmals behandeln und eine Krankheit kann ebenfalls mehrmals behandelt werden. In einer Behandlung können beliebig viele (aber mindestens ein) Medikament angewendet werden. Ein Medikament kann in beliebig vielen Behandlungen Anwendung finden.

Aufgabe 2: Abbildung eines ER-Diagramms auf das relationale Datenmodell

[12 P.]

Gegeben sei folgendes ER-Diagramm:



Entwickeln Sie aus dem dargestellten ER-Diagramm ein entsprechendes relationales Datenbankschema anhand der in der Vorlesung erläuterten Abbildungsregeln. Stellen Sie sicher, dass Ihr Datenbankschema die minimale Anzahl von Relationen aufweist. Stellen Sie das resultierende DB-Schema dar, in dem Sie die notwendigen Relationenschemata in der Form

$$\text{Relationenname}(\text{Attributname}_1, \text{Attributname}_2, \dots, \text{Attributname}_n)$$

anführen und dabei jeweils den Primärschlüssel unterstreichen. Gegebenenfalls enthaltene Fremdschlüssel sind zu 'unterstricheln' und durch die aus den Übungen bekannte Pfeilnotation zu spezifizieren:

$$\underline{\underline{Attr_i}} \rightarrow \underline{\underline{Rel_b.Attr_j}}$$

Hinweis: Die dargestellten (1:n)-Beziehungen haben folgende Semantik:

- (i) Eine Person kann nur in einer Partei Mitglied sein; eine Partei kann mehrere Mitglieder haben.
- (ii) Eine Partei ist Teil höchstens einer Fraktion; einer Fraktion können mehrere Parteien angehören.

Hinweis: Die dargestellte (m:n)-Beziehung 'istMitglied' zwischen den Entity-Typen Abgeordneter und Ausschuss hat folgende Semantik:

- (iv) Ein Abgeordneter kann in einem Ausschuss maximal ein Amt belegen.

Aufgabe 3: Relationenalgebra und SQL

[12 P.]

Gegeben seien die folgenden Relationenschemata:

Person(PNr, Name, EMail, Alter, Werber \rightarrow Person.PNr)

Artikel(ANr, Bezeichnung, Menge, Startpreis, Verkäufer \rightarrow Person.PNr)

Gebot(Bieter \rightarrow Person.PNr, Artikel \rightarrow Artikel.ANr, Zeit, Betrag)

Benutzen Sie zur Lösung der folgenden Aufgaben ausschließlich die in der Vorlesung vorgestellten Operatoren der Relationenalgebra!

[3 P.] a) Geben Sie einen Relationenalgebra-Ausdruck an, der alle Artikel (ANr, Bezeichnung), auf welche eine Person mit dem Namen 'Müller' geboten hat, ausgibt.

[3 P.] b) Geben Sie einen Relationenalgebra-Ausdruck an, der alle Personen (PNr), welche noch keinen Artikel mit der Bezeichnung 'Buch' verkauft haben, ausgibt.

[3 P.] c) Geben Sie eine SQL-Anweisung an, welche die gleiche Ergebnismenge liefert wie der folgende Relationenalgebra-Ausdruck:

$$\pi_{Name}((Person \bowtie_{PNr=Bieter} Gebot \bowtie_{Artikel=ANr} Artikel) \bowtie_{Werber=PNr \wedge Verkäufer=PNr} (\pi_{PNr}(Person)))$$

[3 P.] d) Der kleinste in der Datenbank gespeicherte Startpreis eines Artikels beträgt 10 € und der größte 159 €. Bestimmen Sie die Selektivität der Selektion $\sigma_{Startpreis \geq 46}$ unter der Annahme einer Gleichverteilung:

Aufgabe 4: SQL**[14 P.]**

Gegeben seien die Relationenschemata aus Aufgabe 3:

Person(PNr, Name, EMail, Alter, Werber \rightarrow Person.PNr)Artikel(ANr, Bezeichnung, Menge, Startpreis, Verkäufer \rightarrow Person.PNr)Gebot(Bieter \rightarrow Person.PNr, Artikel \rightarrow Artikel.ANr, Zeit, Betrag)

Formulieren Sie für die in den nachfolgenden Teilaufgaben angeführten, natürlichsprachlich formulierten Anfragen entsprechende SQL-Anweisungen.

Verwenden Sie den in der Vorlesung verwendeten SQL-Standard.

Das SQL-Schlüsselwort JOIN darf dabei nicht verwendet werden.

- a) Die PNr, die EMail und alle gebotenen Beträge der Personen namens 'Harry' nach PNr absteigend sortiert. [2 P.]
- b) Die PNr, der Name sowie der in allen Geboten abgegebene Gesamtbetrag jeder Person, die älter als 30 Jahre alt ist. [3 P.]
- c) Die Namen und EMail aller Personen, die noch nie geboten haben. [3 P.]
- d) Die ANr, die Bezeichnung und den Startpreis jedes Artikels, dessen Verkäufer von einer Person mit dem Namen 'Bond' erworben wurde. [3 P.]
- e) Die PNr und das durchschnittliche Alter der von ihm erworbenen Personen für jeden Werber, der maximal 10 Personen erworben hat. [3 P.]

Aufgabe 5: Änderungsoperationen auf Sichten

[12 P.]

Ausgehend von der Basisrelation

Musiker (PNr, Name, Alter, Genre, Vermögen, Status)

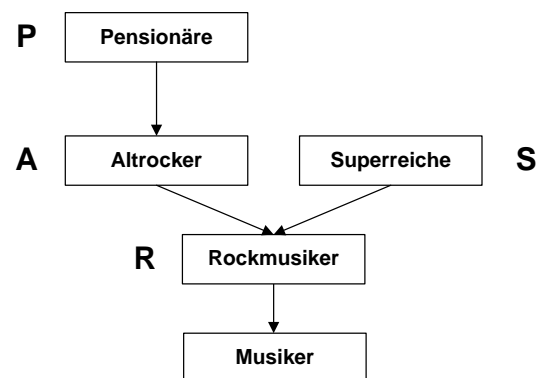
seien folgende Sichten definiert:

```
CREATE VIEW Rockmusiker
AS SELECT * FROM Musiker
WHERE Genre = 'Rock';

CREATE VIEW Altrocker
AS SELECT * FROM Rockmusiker
WHERE Alter > 50;

CREATE VIEW Pensionäre
AS SELECT * FROM Altrocker
WHERE Status = 'Ruhestand';

CREATE VIEW Superreiche
AS SELECT * FROM Rockmusiker
WHERE Vermögen > 100M;
```



Da diese Sichten Änderungsoperationen erlauben, **könnten** sie alle mit dem Zusatz **WITH CASCADED CHECK OPTION** definiert werden. In der Tabelle auf Seite 11 unten sind verschiedene Konfigurationen der Sichten mit bzw. ohne Check-Option dargestellt. Ein Spaltenwert 'CASC' bedeutet dabei, dass die Option **WITH CASCADED CHECK OPTION** für die zugehörige Sicht gesetzt ist; ein Spaltenwert '-' bedeutet, dass die zugehörige Sicht ohne **CHECK OPTION** definiert wurde.

Betrachten Sie folgende Änderungsoperationen:

- UPDATE** Rockmusiker
SET Alter = 49
WHERE Alter = 51
AND Status = 'aktiv'
AND Vermögen < 50M;
- INSERT INTO** Superreiche
VALUES (12, 'Madonna', 50, 'Pop', 500M, 'aktiv');
- UPDATE** Rockmusiker
SET Status = 'aktiv'
WHERE Name = 'Beatles'
AND Alter > 60
AND Vermögen > 200M;

Matrikel-Nr:

Name:

d) **UPDATE** Pensionäre
SET Alter = 48
WHERE Vermögen > 200M;

Markieren Sie im unteren Teil der untenstehenden Tabelle, welche dieser Änderungsoperationen von einem Datenbanksystem in den verschiedenen Konfigurationen erlaubt bzw. zurückgewiesen werden. Die Lösungen zur ersten Änderungsoperation sind als Hilfestellung bereits angegeben. Für die Fälle, in denen die Änderung bzw. das Einfügen zulässig ist, geben Sie in der Tabelle außerdem an, in welchen Sichten **auf jeden Fall alle** geänderten/eingefügten Tupel nach Abschluss der Operation sichtbar werden. Verwenden Sie dabei die folgende Notation:

nein Operation wird für die Konfiguration zurückgewiesen
ja/Sicht-Liste Operation ist für die Konfiguration zulässig.
 Alle geänderten/eingefügten Tupel sind anschließend in
 den in der Sicht-Liste aufgeführten Sichten sichtbar.

Zur Vereinfachung sind anstelle der kompletten Sichtennamen die folgenden Abkürzungen zu verwenden:

Rockmusiker R
Altrock A
Pensionäre P
Superreiche S

Rockmusiker	-	-	CASC	-
Altrock	-	CASC	-	-
Pensionäre	-	-	-	CASC
Superreiche	-	-	CASC	-
a)	ja/R	ja/R	ja/R	ja/R
b)				
c)				
d)				

Aufgabe 6: Optimierung

[6 P.]

Gegeben seien folgende Relationenschemata:

Eigentümer(EID, Vorname, Nachname, Wohnsitz → Immobilie.IID)

Handwerksfirma(HID, Name, Fachgebiet, Firmensitz → Immobilie.IID)

Immobilie(IID, Adresse, Baufirma → Handwerksfirma.HID, Eigentüemer → Eigentüemer.EID)

[3 P.]

1. Für die nachfolgende Anfrage soll eine algebraische Optimierung durchgeführt werden. Optimieren Sie die vorgegebene Anfrage anhand der in der Vorlesung eingeführten Regeln und zeichnen Sie den entsprechenden (optimierten) Operatorbaum. (Projektionen sollen dabei jedoch **nicht** nach unten gezogen werden.)

[3 P.]

2. Bewerten Sie den Operatorbaum mit den Kardinalitäten der Zwischenergebnisse. (Die Anzahl der Attribute soll dabei **nicht** betrachtet werden.) Für die zugehörige Datenbank werden folgende Kardinalitäten angenommen:
 $\text{Card}(\textit{Eigentüemer}) = 50$, $\text{Card}(\textit{Immobilie}) = 1000$
und $\text{Card}(\textit{Handwerksfirma}) = 20$. Unter den Handwerksfirmen gibt es nur 2 verschiedene Fachgebiete wobei von denen jede gleich oft auftritt.

$$\pi_{\text{Name}}(\sigma_{\text{Fachgebiet}=\text{Maurer}}(\sigma_{\text{Wohnsitz}=\text{IID}}(\sigma_{\text{Baufirma}=\text{HID}}(((\sigma_{\text{EID}=1}(\textit{Eigentüemer}) \times \textit{Immobilie}) \cup (\sigma_{\text{EID}=2}(\textit{Eigentüemer}) \times \textit{Immobilie})) \times \textit{Handwerksfirma}))))$$

Aufgabe 7: Transaktionen

[8 P.]

- a) Gegeben sind die drei Objekte x , y und z , welche von den Transaktionen T_1 , T_2 und T_3 gelesen bzw. geschrieben werden. Dabei bezeichnet $w_i(x)$ den Schreibzugriff der Transaktion T_i auf das Objekt x und $r_i(x)$ den Lesezugriff der Transaktion T_i auf x .

[6 P.]

Bestimmen Sie bei den folgenden drei Schedules S_1 , S_2 und S_3 jeweils, ob diese serialisierbar sind oder nicht. Begründen Sie Ihre Antwort, indem Sie entweder alle äquivalenten, seriellen Transaktionsreihenfolgen (z.B. $T_1T_2T_3$) oder einen auftretenden Konfliktzyklus angeben.

- i) $S_1 = r_1(y) \ r_2(z) \ w_3(y) \ w_2(x) \ r_3(y) \ r_1(y) \ r_3(x)$
- ii) $S_2 = r_1(z) \ r_3(x) \ r_2(x) \ r_1(y) \ w_2(z) \ r_3(y) \ w_3(y)$
- iii) $S_3 = w_2(y) \ r_3(z) \ r_1(x) \ r_1(y) \ w_3(x) \ w_3(z)$

- b) **Anomalien bei unkontrolliertem Mehrbenutzerbetrieb**

Betrachten Sie die zwei Transaktionen T_1 und T_2 , deren Operationen ohne Mehrbenutzerkontrolle in der nachfolgend angegebenen Reihenfolge ausgeführt werden. Geben Sie eine Operation von T_2 zum Zeitschritt 3 an (bei Platzmangel bitte unter, **nicht** in die Tabelle schreiben), die für T_1 ein non-repeatable Read verursacht, und begründen Sie kurz die Entstehung dieser Anomalie im vorliegenden Beispiel.

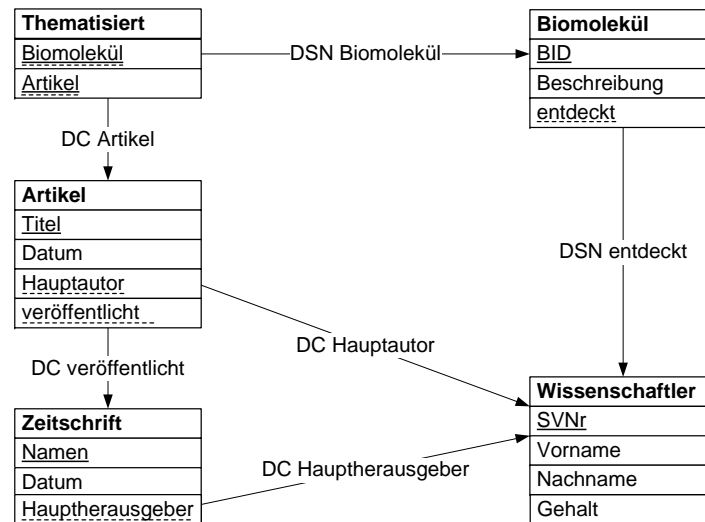
[2 P.]

Zeitschritt	T_1	T_2
1	SELECT SUM(Gehalt) FROM Personal WHERE Abteilung = 'Treadstone';	
2		SELECT AVG(Alter) FROM Personal;
3		
4	SELECT SUM(Gehalt) FROM Personal WHERE Abteilung = 'Treadstone';	
5	COMMIT;	
6		COMMIT;

Aufgabe 8: Referentielle Aktionen

[12 P.]

Gegeben sei folgender Referenzgraph für eine Biomoleküldatenbank.



Die referentiellen Aktionen der Fremdschlüssel in den Abbildungen sind jeweils abgekürzt mit 'DC' für 'ON DELETE CASCADE', 'DR' für 'ON DELETE RESTRICT' bzw. 'DSN' für 'ON DELETE SET NULL'.

[1 P.]

1. Welche Anforderung erfüllt ein (bzgl. der referentiellen Aktionen) sicheres Schema?

[1 P.]

2. Handelt es sich im vorliegenden Fall um ein sicheres Schema?

- ☐ Ja
☐ Nein

3. Die Fremdschlüssel der Relation 'Artikel' wurden durch folgende SQL-DDL-Anweisung erzeugt: [5 P.]

```
CREATE TABLE Artikel (  
    Titel          VARCHAR(50) PRIMARY KEY,  
    Datum          DATE        NOT NULL,  
    Hauptautor     INTEGER      NOT NULL,  
    veröffentlicht VARCHAR(50) NOT NULL,  
    CONSTRAINT fkHauptautor FOREIGN KEY (Hauptautor)  
        REFERENCES Wissenschaftler(SVNr) ON DELETE CASCADE ,  
    CONSTRAINT fkVeröffentlicht FOREIGN KEY (veröffentlicht)  
        REFERENCES Zeitschrift(Name) ON DELETE CASCADE  
);
```

Für den Fremdschlüssel 'fkVeröffentlicht' aus der Relation 'Artikel' soll nun die referentielle Aktion von 'ON DELETE CASCADE' auf „ON DELETE RESTRICT“ geändert werden:

Geben Sie eine Folge von **SQL-DDL**-Anweisungen an, welche diese Schemaänderung realisiert und dabei die Löschung der Relation selbst vermeidet.

4. Handelt es sich bei dem geänderten Schema um ein sicheres Schema? [5 P.]
Falls das Schema nicht sicher ist, beschreiben Sie ein Szenario, in dem das Ergebnis einer Änderungs-/Löschoperation von der Auswertungsreihenfolge der referentiellen Aktionen abhängig ist.

- ☐ Ja
☐ Nein

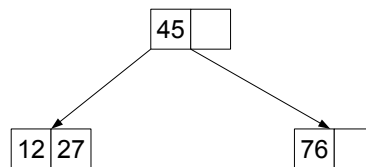
Aufgabe 9: B-Bäume

[11 P.]

Anmerkung: Die B-Bäume sind im Folgenden stets gemäß der vereinfachten Darstellungsart aus der Vorlesung abgebildet.

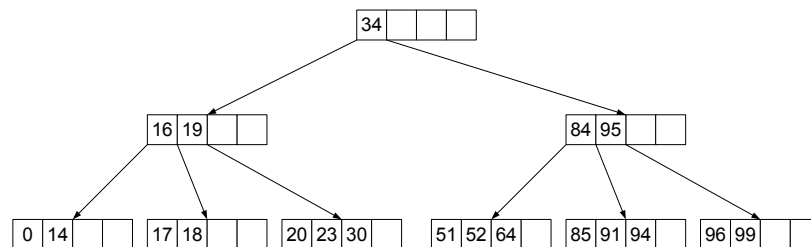
[5 P.]

- a) Nehmen Sie den (Standard-) Split-Faktor 1 an und fügen Sie in den unten abgebildeten **B-Baum** der Klasse $\tau(1, h)$ die Datensätze mit den Schlüsselwerten **3, 17, 83, 33 und 79** in dieser Reihenfolge ein. Zeichnen Sie den Baum nach jedem Split-Vorgang und erläutern Sie kurz die durchgeführten Maßnahmen (Splitten, einfaches Einfügen).



[6 P.]

- b) Löschen Sie aus dem unten abgebildeten **B-Baum** der Klasse $\tau(2, h)$ die Datensätze mit den Schlüsselwerten **17, 96, 18 und 16** (in dieser Reihenfolge). Zeichnen Sie den Baum nach jedem Löschvorgang und geben Sie jeweils kurz an, welche konkrete Maßnahme Sie durchgeführt haben (Mischen, Ausgleichen, einfaches Löschen). Für Ausgleichs- und Mischoperationen sollen nur direkt benachbarte Geschwisterknoten (bevorzugt der rechte) herangezogen werden.



Aufgabe 10: Berechnungen in B- und B*-Bäumen

[15 P.]

- a) Gegeben seien ein maximal gefüllter B-Baum der Klasse $\tau(3, 2)$ und ein minimal gefüllter B*-Baum der Klasse $\tau(2, 3, 3)$. [11 P.]

Anmerkung: Begründen Sie in den folgenden Teilaufgaben Ihr jeweiliges Ergebnis durch einen Rechenweg.

- (1) Wie groß ist die Differenz der Anzahl an Datensätzen der jeweiligen Bäume? [6 P.]
 - (2) Wenn man sequentiell alle Datensätze aus einem der beiden Bäume lesen möchte, wieviele verschiedene Knoten (Seiten) muss man in dem jeweiligen Baum lesen? [3 P.]
 - (3) Wenn man wahlfrei einen Datensatz aus einem der beiden Bäume lesen möchte, wieviel Knoten (Seiten) muss man im jeweiligen Baum maximal lesen? [2 P.]
- b) Gegeben sei ein **B*-Baum** der Klasse $\tau(k, k^*, h)$. Innerhalb des zugehörigen (fiktiven) DBS wird vereinfachend eine Seitengröße von 800B ($L=800B$) angenommen. Die weiteren Kenngrößen sind dabei wie folgt festgelegt: [4 P.]

$$\begin{aligned}
 l_M &= 4B \\
 l_D &= 160B \\
 l_K &= 5B \\
 l_P &= 6B
 \end{aligned}$$

Die inneren Baum- und Blattknoten entsprechen jeweils genau einer Seite. Wie groß sind demnach k und k^* ?

Begründen Sie Ihr Ergebnis durch einen Rechenweg.

Hinweis: Die oben genannten Kenngrößen haben folgende Bedeutung:

l_M	die Länge der Seiteninformationen einer Seite
l_D	die Länge eines Datensatzes
l_K	die Länge eines Schlüsselwertes
l_P	die Länge eines Zeigers