



Universität Hamburg
Fachbereich Informatik
Datenbanken und Informationssysteme
Prof. Dr. N. Ritter

Modul
Grundlagen von Datenbanken
Wintersemester 2012/2013

Probeklausur



Aufgabe	1	2	3	4	5	6	7	8	9	10	Gesamt
Punkte	18	12	12	14	12	6	8	11	12	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.
- Überprüfen Sie sofort nach Erhalt der Klausur die Vollständigkeit der Unterlagen (26 Seiten).
- Bitte lassen Sie die Klausur zusammengeheftet. Notieren Sie alle Lösungen direkt auf den Aufgabenblättern.
- Verwenden Sie kein eigenes Papier, sondern nur die zur Verfügung gestellten Blätter.
- Bitte schreiben Sie mit blauem oder schwarzem Schreiber - Bleistift und Rotstift sind nicht erlaubt.
- Schreiben Sie auf jedes Blatt Ihren Namen und Ihre Matrikelnummer.
- Es sind keine Hilfsmittel zugelassen.
- Bitte legen Sie Ihren Studentenausweis und einen Lichtbildausweis auf den Tisch.
- Schalten Sie Ihr Mobiltelefon aus!

Viel Erfolg!

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 notieren Sie die Abbildungstypen in der Form 1:1, 1:n, n:m und verfeinern Sie diese durch Kardinalitätsrestriktionen (Notation: [min,max]), falls notwendig. 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).

(Anmerkung: Etwaige (Existenz-)Abhängigkeiten sollen nicht modelliert werden.)

Es wird ein Ärztehaus, dessen Angestellte und Patienten sowie deren Behandlungen betrachtet. Die Angestellten des Ärztehauses 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. Einige Hilfskräfte können (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 Behandlung – welche durch eine eindeutige Behandlungsnummer (BehNr) gekennzeichnet ist – eines Patienten aufgrund einer bestimmten Krankheit erfolgt durch genau einen Arzt, mit einem oder mehreren Medikamenten und findet an einem festen Behandlungsdatum statt. Ein Patient kann, auch wegen derselben Krankheit, mehrmals behandelt werden und ein Arzt kann mehrere Behandlungen durchführen. Dabei kann ein Patient bzgl. einer Krankheit auch mehrmals von demselben Arzt behandelt werden. 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.

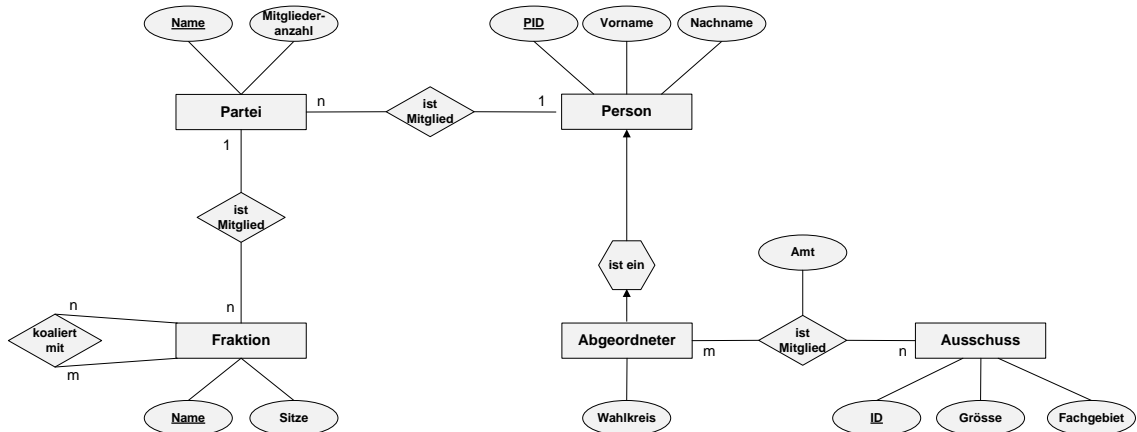
Matrikel-Nr:

Name:

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:

$$\underline{\text{Attr}_i} \rightarrow \text{Rel}_b.\underline{\text{Attr}_j}$$

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.

Matrikel-Nr:

Name:

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.

Matrikel-Nr:

Name:

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

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

- d) Der kleinste in der Datenbank gespeicherte Startpreis eines Artikels [3 P.]
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.

[2 P.]

- a) Die PNr, die EMail und alle gebotenen Beträge der Personen namens 'Harry' nach PNr absteigend sortiert.

[3 P.]

- b) Die PNr, der Name sowie der in allen Geboten abgegebene Gesamtbeitrag jeder Person, die älter als 30 Jahre alt ist.

Matrikel-Nr:

Name:

c) Die Namen und EMail's 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)

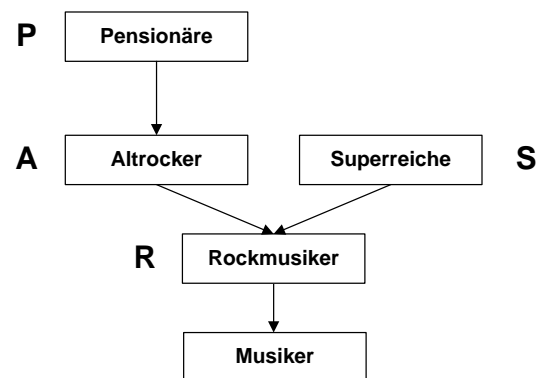
seien folgende Sichten definiert:

```
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
Altrockers A
Pensionäre P
Superreiche S

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

Aufgabe 6: Normalformenlehre

[6 P.]

Gegeben ist die Relation R mit den Attributen N, O, D, T, Z und S, sowie der Menge F an funktionalen Abhängigkeiten

$$F = \{FA_1, FA_2, FA_3, FA_4, FA_5\}.$$

Die Wertebereiche der Attribute sind alle atomar.

$$R(N, O, D, T, Z, S)$$

$$FA_1 = N \rightarrow S$$

$$FA_2 = T \rightarrow O, Z$$

$$FA_3 = N, D \rightarrow T$$

$$FA_4 = S, D \rightarrow N$$

$$FA_5 = Z \rightarrow O$$

[2 P.]

i) Bestimmen Sie die Schlüsselkandidaten von R bezüglich F.

[1 P.]

ii) Bestimmen Sie die Nicht-Primärattribute (Nicht-Schlüsselattribute) von R bezüglich F.

Matrikel-Nr:

Name:

- iii) Nehmen Sie an, dass einer der in Aufgabenteil i) ermittelnden Schlüsselkandidaten als Primärschlüssel verwendet wird. In welchen Normalformen befindet sich das Relationenschema R bezüglich F ? Begründen Sie Ihre Antwort, indem Sie darlegen, warum sich das Relationenschema in genau diesen Normalformen befindet und warum die anderen Normalformen nicht vorliegen. [3 P.]

(Anmerkung: Betrachten Sie dabei lediglich die 1., 2. und 3. Normalform.)

Aufgabe 7: Transaktionen

[8 P.]

[6 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 .

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) $S_3 = w_2(y) \ r_3(z) \ r_1(x) \ r_1(y) \ w_3(x) \ w_3(z)$

Matrikel-Nr:

Name:

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;

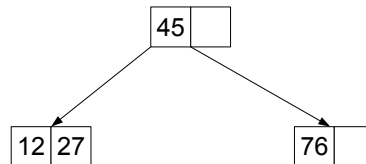
Aufgabe 8: B-Bäume

[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).

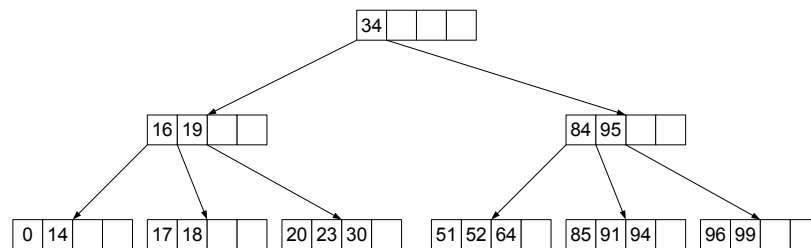


Matrikel-Nr:

Name:

[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.



Matrikel-Nr:

Name:

Aufgabe 9: Berechnungen in B- und B*-Bäumen

[12 P.]

[8 P.]

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

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

[5 P.]

- (1) Wie groß ist die Differenz der Anzahl an Datensätzen der jeweiligen Bäume?

[2 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) Wenn man wahlfrei einen Datensatz aus einem der beiden Bäume lesen möchte, wieviel Knoten (Seiten) muss man im jeweiligen Baum maximal lesen? [1 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

Aufgabe 10: XPath

[15 P.]

Gegeben sei nachfolgende DTD (Sie beschreibt die Struktur einer Ansammlung an Vereinsligen).

```
<?xml version="1.0" encoding="UTF-8"?>

<!ELEMENT Liga (Sportart, Präsidenten, Verein+)>
<!ATTLIST Liga
    LNr ID #REQUIRED>

<!ELEMENT Sportart (#PCDATA)>

<!ELEMENT Präsident (#PCDATA)>
<!ATTLIST Präsident
    PNr IDREF #REQUIRED>

<!ELEMENT Verein (Vorstand, Trainer, Spieler+)>
<!ATTLIST Verein
    Name ID #REQUIRED>

<!ELEMENT Vorstand (Präsident, Manager)>

<!ELEMENT Manager (#PCDATA)>
<!ATTLIST Manager
    PNr IDREF #REQUIRED>

<!ELEMENT Trainer (#PCDATA)>
<!ATTLIST Trainer
    PNr IDREF #REQUIRED>

<!ELEMENT Spieler (Nummer, Position)>
<!ATTLIST Spieler
    PNr IDREF #REQUIRED>

<!ELEMENT Nummer (#PCDATA)>
<!ELEMENT Position (#PCDATA)>

<!ELEMENT Person (Vorname, Name, Alter)>
<!ATTLIST Person
    PNr ID #REQUIRED>

<!ELEMENT Vorname (#PCDATA)>
<!ELEMENT Name (#PCDATA)>
<!ELEMENT Alter (#PCDATA)>
```


Matrikel-Nr:

Name:

- a) Geben Sie für die nachfolgenden XPath-Ausdrücke jeweils eine natürlichsprachliche Anfrage an.

i) `//Verein[Name/text()=„FC St.Pauli“]/Spieler[Position=`
`(//Spieler[fn:id(@PNr)/Name/text()=„Klose“]/Position)]/Nummer` [3 P.]

ii) `fn:id(/Liga[Sportart/text()=„Handball“]`
`/Verein[Name/text()=„HSV“]/Vorstand/Präsident/@PNr)/Alter` [3 P.]

[3 P.]

- iii) `fn:id(//Präsidenten[@PNr=(//Person[Vorname/text()='Klaus']/@PNr)]
/parent::Vorstand/Manager/@PNr)/element()`

- b) Geben Sie zur Ermittlung der nachfolgenden Informationen jeweils einen XPath-Ausdruck an.

[2 P.]

- i) Die Positionen der Spieler, welche bei einem Verein mit dem Namen „HSV“ spielen.

Matrikel-Nr:

Name:

ii) Die Namen der Vereine, deren Manager jünger als 60 Jahre alt ist. [2 P.]

iii) Die Namen der Personen, die für einen Verein mit dem Namen „Telekom Baskets Bonn“ in einer Basketball Liga spielen. [2 P.]

Matrikel-Nr:

Name:
