	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>		WS 2018/19
	Aufgabenzettel	<b>3 (Lösungsvorschläge)</b>		
	Gesamtpunktzahl	<b>40</b>		
	Ausgabe	<b>Mi. 14.11.2018</b>	Abgabe	<b>Fr. 30.11.2018</b>

# 1 Informationsmodellierung

[17 P.]

Erfassen Sie die im Folgenden beschriebenen Informationsstrukturen (Ausschnitt einer Buch-Verwaltung) 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 Beziehungskardinalitäten in der Form [min;max]. Verwenden Sie unbedingt die aus der Vorlesung bekannte Notation. Andere Notationen werden nicht gewertet! Ausnahme: Die Leserichtung der Beziehungskardinalitäten bleibt Ihnen überlassen, muss aber eindeutig als solche markiert werden (z.B. durch ein ausformuliertes Beispiel). Benutzen Sie möglichst wenige Entitäten (Ausnahme: Vererbung). Modellieren Sie Abhängigkeiten zwischen Entitäten nur, wenn diese im Text explizit (z.B. durch die Angabe einer schwachen Entität) beschrieben sind.

*Es gibt Personen, Bücher, Verlage, Preisverleihungen und Selbsthilfegruppen gegen Schreibblockaden (kurz SHG). Eine Person hat eine eindeutige Sozialversicherungsnummer (kurz SVNr), einen Vornamen, einen Nachnamen und ein Geburtsdatum. Die Menge der Personen lässt sich in drei Teilmengen untergliedern: den Schriftstellern, den Verlegern und den Lektoren.*

*Ein Schriftsteller kann Mitglied bei keiner, einer oder mehreren Selbsthilfegruppen sein. Eine SHG besitzt einen eindeutigen Namen und einen Versammlungsort. Bei einer SHG müssen mindestens 5 und maximal 20 Schriftsteller Mitglied sein.*

*Schriftsteller sind Autoren von Büchern, wobei jeder Schriftsteller mindestens ein Buch geschrieben haben muss, aber auch Autor beliebig vieler weiterer Bücher sein kann. Ein Buch hat eine eindeutige ISBN, einen Titel und ein Erscheinungsjahr. Ein Buch kann einen oder mehrere Autoren besitzen und wird von genau einem Verlag verlegt.*

*Ein Verlag hat einen eindeutigen Namen und eine Adresse, die sich aus Straße, Hausnummer und PLZ zusammensetzt. Ein Verlag kann kein, ein oder mehrere Bücher verlegen. Bei einem Verlag können einer oder mehrere Verleger arbeiten. Ein Verleger arbeitet für genau einen Verlag, und hat eine Abteilung in der er tätig ist.*

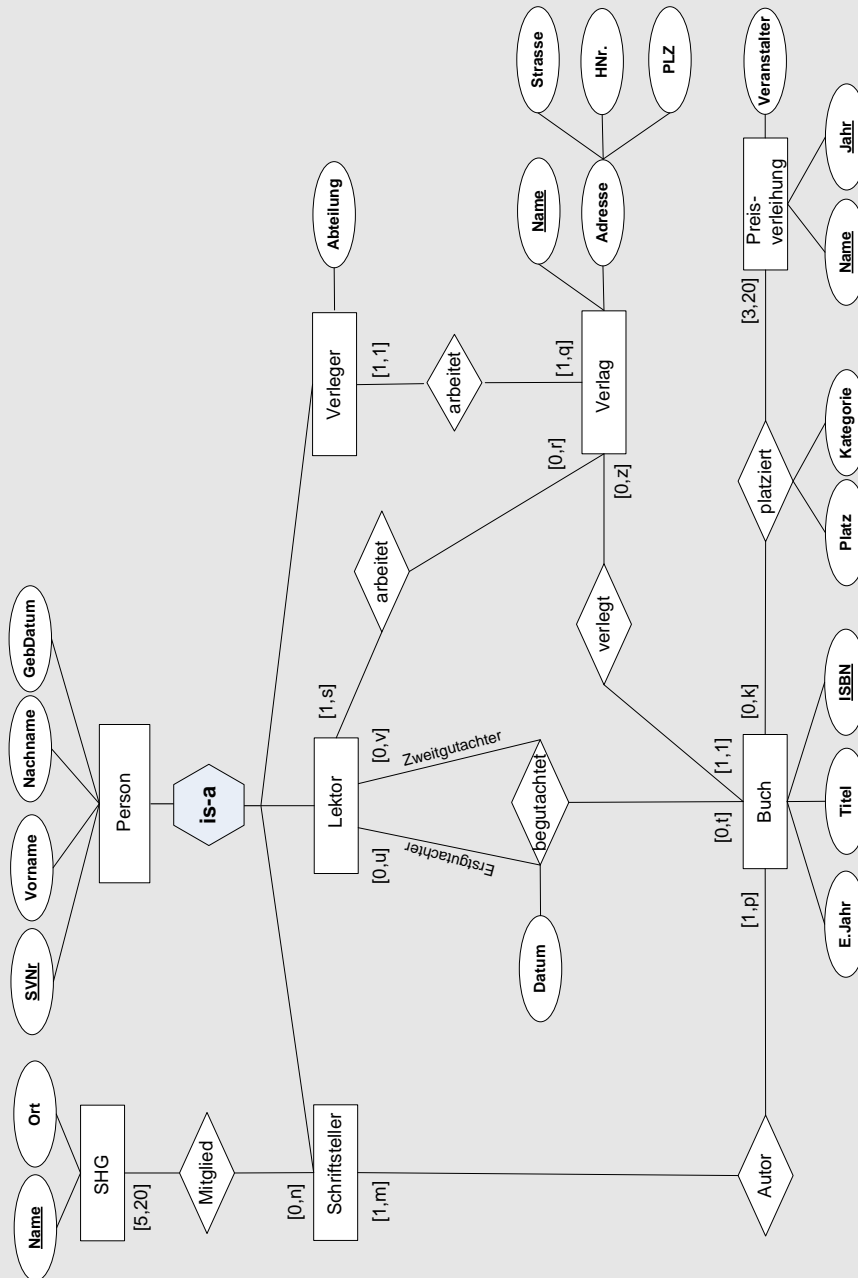
*Ein Lektor kann für einen oder mehrere Verläge arbeiten. Für einen Verlag können beliebig viele Lektoren arbeiten.*

*Ein Buch kann keinmal, einmal oder mehrmals begutachtet werden, wobei jede Begutachtung mit einem Datum versehen ist. An jeder Begutachtung sind genau ein Buch und genau zwei Lektoren beteiligt. Einer der Lektoren übernimmt dabei die Rolle des Erstgutachters und der andere die Rolle des Zweitgutachters. Ein Lektor begutachtet kein, ein oder mehrere Bücher (egal ob als Erst- oder als Zweitgutachter).*

*Ein Buch kann sich bei keiner, einer oder mehreren Preisverleihungen platzieren, wobei jede Platzierung Informationen über den Platz und die Kategorie beinhaltet. Eine Preisverleihung hat einen Namen, ein Jahr und einen Veranstalter, wobei die Kombination aus Name und Jahr eindeutig ist. Bei einer Preisverleihung sind mindestens 3 und maximal 20 Bücher platziert.*

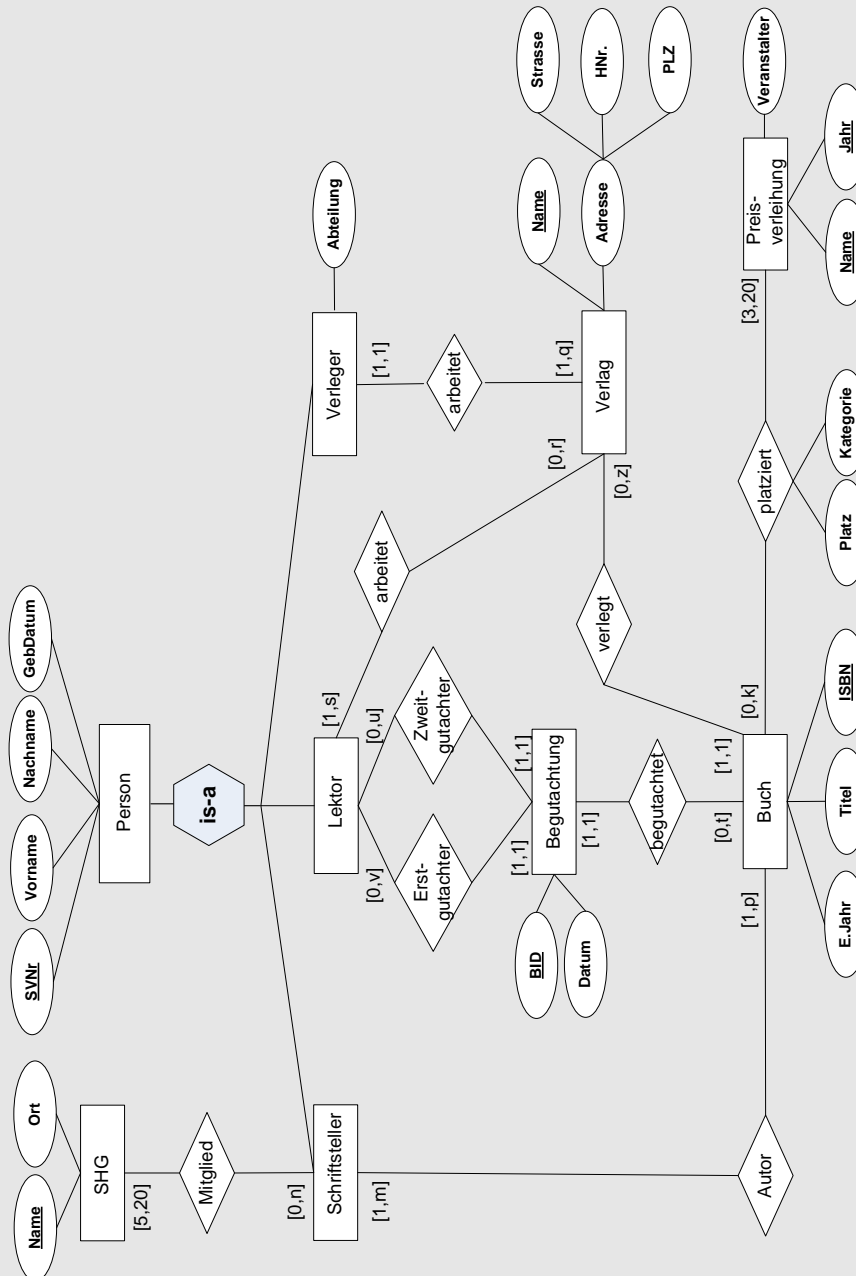



Lösungsvorschlag:





Lösungsvorschlag:

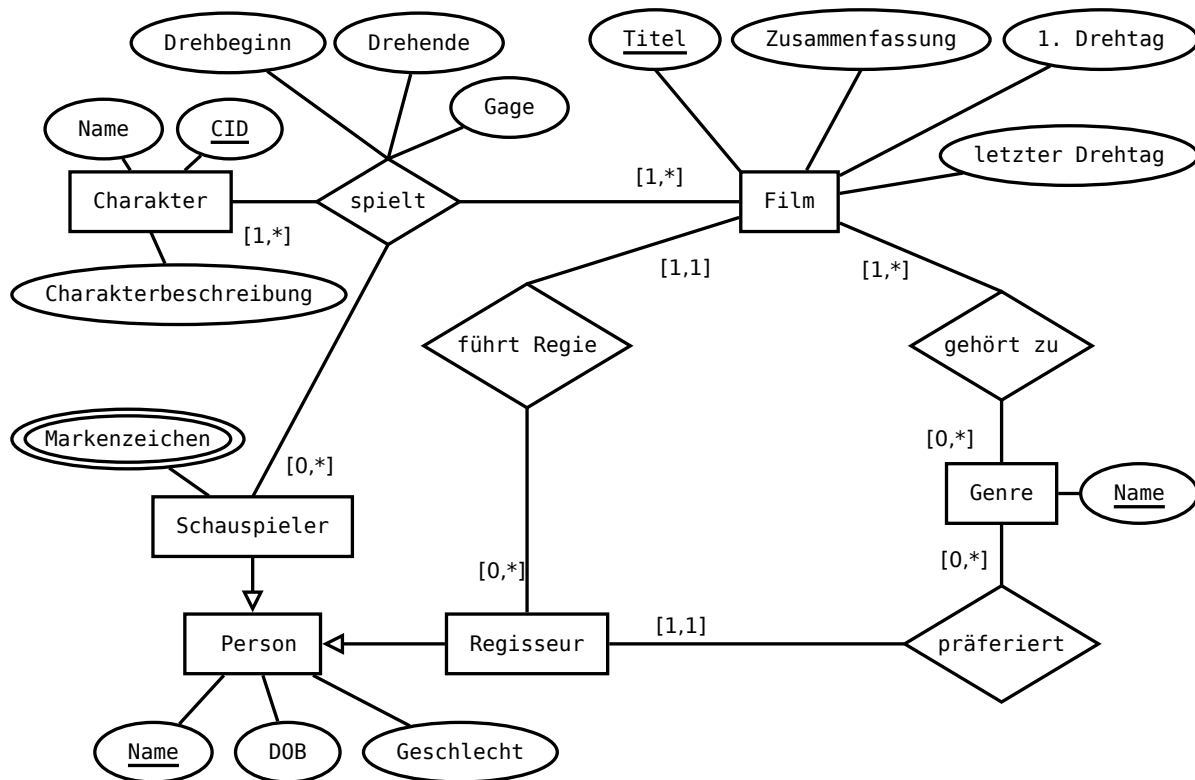


	Lehrveranstaltung	Grundlagen von Datenbanken		WS 2018/19
	Aufgabenzettel	3 (Lösungsvorschläge)		
	Gesamtpunktzahl	40		
	Ausgabe	Mi. 14.11.2018	Abgabe	Fr. 30.11.2018

## 2 Logischer Entwurf

[10 P.]

Gegeben sei folgendes ER-Diagramm:




Vollziehen Sie einen logischen Entwurf und transformieren Sie das unten abgebildete ER-Diagramm in ein relationales Datenbankschema. Versuchen Sie das Datenbankschema dabei möglichst minimal zu halten, ohne dadurch Informationseinbußen hinnehmen zu müssen. Verwenden Sie bitte folgende Schreibweise für eine Relation:

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

Die Attribute eines Primärschlüssels sind mit einer einzigen durchgezogenen Linie zu unterstreichen. Das Gleiche gilt für zusammengesetzte Fremdschlüssel, in diesem Fall ist jedoch eine gestrichelte Linie zu verwenden. Die Referenzen der Fremdschlüssel sollen wie folgt dargestellt werden:

$$\text{Attributname}_i \rightarrow \text{Relationenname}_j.\text{Attributname}_k$$

Hinweis zur Semantik von binären Beziehungen: Nach dem dargestellten ER-Diagramm kann jeder Regisseur bei beliebig vielen Filmen die Regie führen, während jeder Film genau einen Regisseur hat. Die Semantik aller anderen binären Beziehungen ist entsprechend.

	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>		WS 2018/19
	Aufgabenzettel	<b>3 (Lösungsvorschläge)</b>		
	Gesamtpunktzahl	<b>40</b>		
	Ausgabe	<b>Mi. 14.11.2018</b>	Abgabe	<b>Fr. 30.11.2018</b>

### Lösungsvorschlag:

*Genre*(Name)

*Person*(Name, *DOB*, *Geschlecht*)

*Schauspieler*(Name → Person.Name)

*Markenzeichen*(Name → Schauspieler.Name, *Markenzeichen*)

*Regisseur*(Name → Person.Name, *Lieblingsgenre* → Genre.Name)

*Film*(Titel, *Zusammenfassung*, *ersterDrehtag*, *letzterDrehtag*,  
Regisseur → Regisseur.Name)

*Charakter*(CID, *Name*, *Charakterbeschreibung*)

*spielt*(Name → Schauspieler.Name, *Charakter* → Charakter.CID,  
Film → Film.Titel, *Drehbeginn*, *Drehende*, *Gage*)


*gehörtZu*(Film → Film.Titel, *Genre* → Genre.Name)

Alternativ die Modellierung mit dem Hausklassenmodell:

*Person*(Name, *DOB*, *Geschlecht*)

*Schauspieler*(Name, *DOB*, *Geschlecht*)

*Regisseur*(Name, *DOB*, *Geschlecht*, *Lieblingsgenre* → Genre.Name)

	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>		WS 2018/19
	Aufgabenzettel	<b>3 (Lösungsvorschläge)</b>		
	Gesamtpunktzahl	<b>40</b>		
	Ausgabe	<b>Mi. 14.11.2018</b>	Abgabe	<b>Fr. 30.11.2018</b>

### 3 Relationale Algebra und SQL

[9 P.]

Betrachten Sie das folgende relationale Datenbankschema.

<u>Rennfahrer</u>	<u>RID</u>	Vorname	Nachname	Geburt	Geburtsort	<u>Rennstall</u>
	4	Sebastian	Vettel	1987-07-03	Heppenheim (Deutschland)	5
	6	Fernando	Alonso	1981-07-29	Oviedo (Spanien)	32
	8	Max	Verstappen	1997-09-30	Hasselt (Belgien)	2
	9	Lewis	Hamilton	1985-01-07	Stevenage (UK)	31
	20	Valtteri	Bottas	1989-08-28	Nastola (Finnland)	31
	21	Daniel	Riccardo	1989-07-01	Perth (Australien)	2
	44	Kimi	Räikkönen	1979-10-17	Espoo (Finnland)	5


Rennstall → Rennstall.RSID

<u>Rennstall</u>	<u>RSID</u>	Name	Teamchef	Budget
	2	Red Bull	Christian Horner	370
	5	Ferrari	Maurizio Arrivabene	350
	31	Mercedes	Toto Wolf	360
	32	McLaren	Zack Brown	220

<u>Rennort</u>	<u>OID</u>	Name	Strecke
	4	Australien GP	Albert Park Circuit
	15	Bahrain GP	Bahrain International Circuit
	21	Monaco GP	Circuit de Monaco

<u>Platzierung</u>	<u>RID</u>	<u>OID</u>	Platz
	8	4	6
	4	15	1
	20	15	2
	4	4	1
	9	15	3
	21	21	1
	9	4	2
	21	4	4
	4	21	2
	20	4	8
	44	4	3

RID → Rennfahrer.RID, OID → Rennorte.OID

	Lehrveranstaltung	Grundlagen von Datenbanken		WS 2018/19
	Aufgabenzettel	3 (Lösungsvorschläge)		
	Gesamtpunktzahl	40		
	Ausgabe	Mi. 14.11.2018	Abgabe	Fr. 30.11.2018

a) Interpretieren Sie die folgenden relationalen Ausdrücke, indem Sie eine umgangssprachliche Beschreibung sowie die Ergebnisrelation angeben. [3 P.]

i)  $\pi_{Nachname}(Rennfahrer \bowtie \sigma_{Platz=2}(Platzierung) \bowtie \sigma_{Strecke="Circuit de Monaco"}(Rennort))$

**Lösungsvorschlag:**

Den Nachnamen des Rennfahrers, der den zweiten Platz auf der Strecke Circuit de Monaco geholt hat.  
= {(Vettel)}

ii)  $\pi_{Rennstall.Name}(Rennstall \bowtie_{RSID=Rennstall} Rennfahrer \bowtie \pi_{RID}(\sigma_{Name="Australien GP"}(Rennort) \bowtie Platzierung))$

**Lösungsvorschlag:**

Die Namen aller Rennställe, von denen im Rennen in Australien mindestens ein Fahrer platziert war.  
= {(Red Bull), (Ferrari), (Mercedes)}

iii)  $\pi_{Nachname}(\sigma_{Teamchef="Christian Horner"}(Rennstall) \bowtie_{RSID=Rennstall} Rennfahrer)$

**Lösungsvorschlag:**

Die Nachnamen der Fahrer, die für das Team von Christian Horner fahren.  
= {(Verstappen), (Ricardo)}

b) Übersetzen Sie die folgenden umgangssprachlich formulierten Anfragen in einen zugehörigen Ausdruck der relationalen Algebra. Werten Sie die Ausdrücke aus und geben Sie jeweils die Ergebnisrelation an. [4 P.]

i) Namen der Rennställe, in denen junge Fahrer (Geboren in oder nach 1985) angestellt sind


**Lösungsvorschlag:**

$\pi_{Rennstall.Name}(\sigma_{Geburt \geq "1985-01-01"}(Rennfahrer) \bowtie_{Rennstall=RSID} Rennstall)$   
= {(Red Bull), (Mercedes), (Ferrari)}

ii) Persönliche Daten (Vor- und Nachname, Geburtsdatum) aller beim *Australien GP* platzierten Rennfahrer, die bei *Red Bull* angestellt sind.

**Lösungsvorschlag:**

$\pi_{Vorname, Nachname, Geburt}(\sigma_{Name="Red Bull"}(Rennstall) \bowtie_{RSID=Rennstall} Rennfahrer \bowtie Platzierung \bowtie \sigma_{Name="Australien GP"}(Rennort))$   
= {(Max, Verstappen, 1997-09-30), (Daniel, Riccardo, 1989-07-01)}

	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>		WS 2018/19
	Aufgabenzettel	<b>3 (Lösungsvorschläge)</b>		
	Gesamtpunktzahl	<b>40</b>		
	Ausgabe	<b>Mi. 14.11.2018</b>	Abgabe	<b>Fr. 30.11.2018</b>

iii) Alle Informationen der Rennfahrer, die bisher keine Platzierung erreicht haben.

**Lösungsvorschlag:**

$$\begin{aligned} & \pi_{\text{Rennfahrer}} \text{ - } \pi_{\text{Rennfahrer.*}}(\text{Rennfahrer} \bowtie \text{Platzierung}) \\ & = (\pi_{\text{RID}}(\text{Rennfahrer}) - \pi_{\text{RID}}(\text{Platzierung})) \bowtie \text{Rennfahrer} \\ & = \{(6, \text{Fernando, Alonso, 1981-07-29, Oviedo (Spanien)}, 32)\} \end{aligned}$$

iv) Vor- und Nachnamen der Rennfahrerkollegen (selber Rennstall) von einem Rennfahrer mit dem Nachnamen "Vettel", die selber nicht so heißen.

**Lösungsvorschlag:**

$$\begin{aligned} & \pi_{\text{Vorname, Nachname}}(\sigma_{\text{Nachname} \neq \text{"Vettel"}}(\text{Rennfahrer} \bowtie \pi_{\text{Rennstall}}(\sigma_{\text{Nachname} = \text{"Vettel"}}(\text{Rennfahrer})))) \\ & = \{(\text{Kimi, Räikkönen})\} \end{aligned}$$


c) Übersetzen Sie die folgenden umgangssprachlich formulierten Anfragen in einen zugehörigen SQL-Ausdruck. (Verwenden Sie hierfür nicht das Schlüsselwort JOIN.) [2 P.]

i) Persönliche Daten (Vor- und Nachname, Geburtsdatum) aller beim *Australien GP* platzierten Rennfahrer, die bei *RedBull* angestellt sind.

**Lösungsvorschlag:**

```
SELECT
  r.Vorname,
  r.Nachname,
  r.Geburt
FROM
  Rennstall rs,
  Rennfahrer r,
  Platzierung p,
  Rennort ro
WHERE
  rs.RSID = r.Rennstall AND
  r.RID = p.RID AND
  p.OID = ro.OID AND
  rs.Name = "RedBull" AND
  ro.Name = "Australien GP"
```



	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>		WS 2018/19
	Aufgabenzettel	<b>3 (Lösungsvorschläge)</b>		
	Gesamtpunktzahl	<b>40</b>		
	Ausgabe	<b>Mi. 14.11.2018</b>	Abgabe	<b>Fr. 30.11.2018</b>

- ii) Vor- und Nachnamen der Rennfahrerkollegen (selber Rennstall) von einem Rennfahrer mit dem Nachnamen "Vettel", die selber nicht so heißen.

**Lösungsvorschlag:**

```

SELECT
  kollegen.Vorname,
  kollegen.Nachname
FROM
  Rennfahrer sv,
  Rennfahrer kollegen
WHERE
  sv.Nachname = "Vettel" AND
  sv.Rennstall = kollegen.Rennstall AND
  kollegen.Nachname != "Vettel"

```

## 4 Algebraische Optimierung


[4 P.]

Betrachten Sie erneut das Datenbankschema aus Aufgabe 3. In der folgenden Aufgabe sind zwei relationale Ausdrücke angegeben. Beide Ausdrücke liefern dasselbe Ergebnis zurück und sind daher semantisch äquivalent, unterscheiden sich jedoch in ihrem Optimierungsgrad. Zeichnen Sie zu jedem relationalen Ausdruck einen Operatorbaum und bestimmen Sie, welcher der zwei Operatorbäume den höheren Optimierungsgrad besitzt. Begründen Sie Ihre Entscheidung mit Hilfe der in der Vorlesung behandelten Optimierungsheuristiken I-VII (Kapitel 4.).

Für die zugehörige Datenbank werden in dieser Aufgabe folgende Kardinalitäten angenommen:

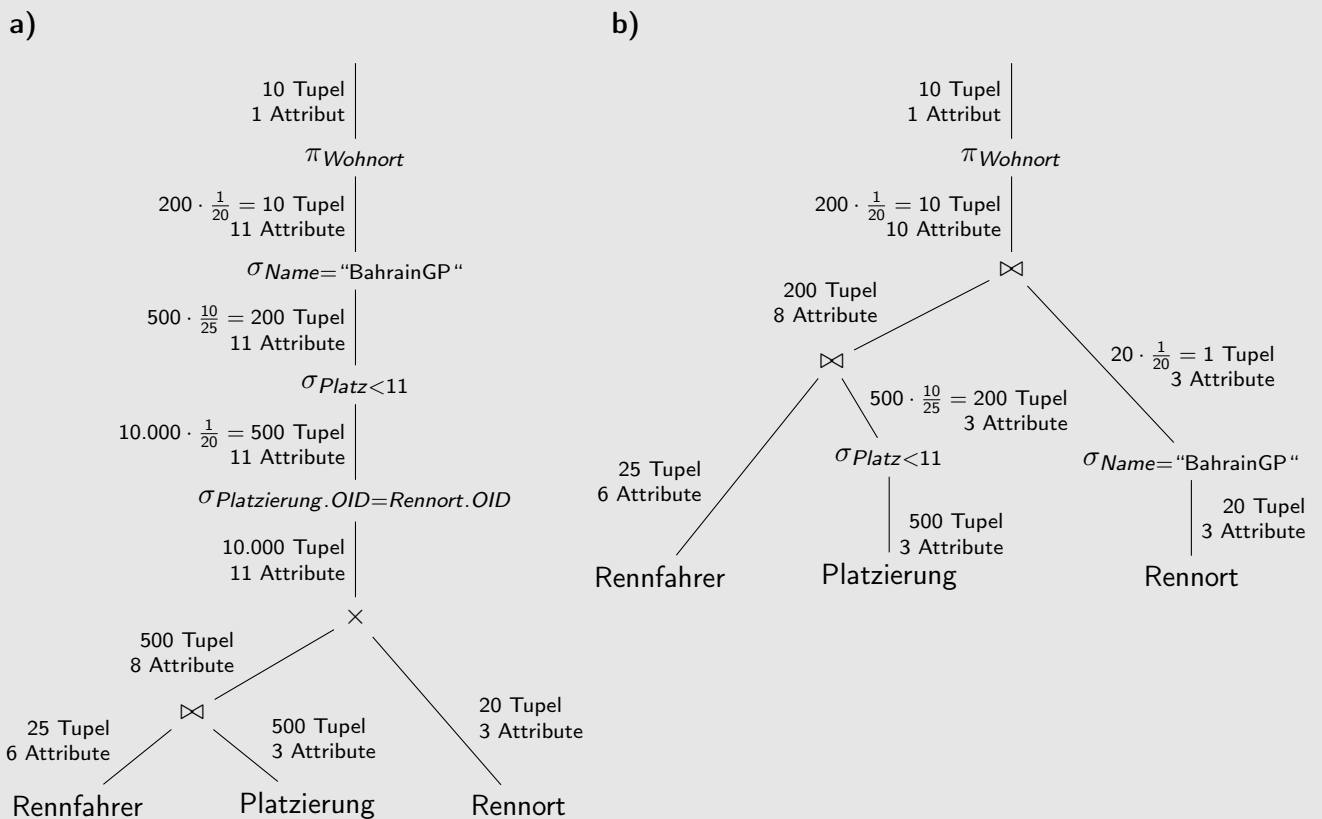
$Card(Rennfahrer) = 25$ ,  $Card(Platzierung) = 500$ ,  $Card(Rennort) = 20$

Es gibt 20 verschiedene Werte für das Attribut Rennort. Bei jedem Rennen sind 25 Rennfahrer platziert. Die Namen von Rennenorten sind eindeutig.

	Lehrveranstaltung	Grundlagen von Datenbanken		WS 2018/19
	Aufgabenzettel	3 (Lösungsvorschläge)		
	Gesamtpunktzahl	40		
	Ausgabe	Mi. 14.11.2018	Abgabe	Fr. 30.11.2018

- a)  $\pi_{Wohnort}(\sigma_{Name="BahrainGP"}(\sigma_{Platz < 11}(\sigma_{Platzierung.OID=Rennort.OID}(((Rennfahrer \bowtie Platzierung) \times Rennort))))$
- b)  $\pi_{Wohnort}((Rennfahrer \bowtie \sigma_{Platz < 11}(Platzierung)) \bowtie \sigma_{Name="BahrainGP"}(Rennort))$

### Lösungsvorschlag:



Der zweite Operatorbaum besitzt den höheren Optimierungsgrad, da hier die Selektionen so früh wie möglich ausgeführt werden. Der Verbund zwischen den Relationen Rennfahrer und Platzierung führt zu einem Zwischenergebnis von etwa 200 Tupeln (vgl. 500 Tupel) und der nachfolgende Verbund mit der Rennort-Relation liefert 10 Tupel (vgl. 500 Tupel).