	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>			WS 2013/14
	Aufgabenzettel	<b>3</b>			
	Gesamtpunktzahl	<b>40</b>			
	Ausgabe	<b>Mi. 13.11.2013</b>	Abgabe	<b>Do. 28.11.2013</b>	


## Aufgabe 1: Konzeptioneller Entwurf

[17 P.]

Erfassen Sie die im Folgenden beschriebenen Informationsstrukturen (Datenbank zur Speicherung von Biomolekülen) 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]. Benutzen Sie möglichst wenige Entity-Typen (Ausnahme: Vererbung). Modellieren Sie Abhängigkeiten zwischen Entitäten nur, wenn diese im Text explizit (z.B. durch die Angabe einer schwachen Entität) beschrieben sind.

*Modelliert werden Biomoleküle. Ein Biomolekül ist entweder ein DNA-Molekül oder ein mRNA-Molekül. Jedes Biomolekül hat eine eindeutige Molekül-ID und eine Beschreibung. Ein Biomolekül kommt in beliebig vielen Organismen vor, wobei jeder Organismus aus mindestens einem Biomolekül besteht. Ein Organismus wird eindeutig über seine Taxonomie-ID identifiziert und hat einen Namen sowie einen Trivialnamen. Ein Biomolekül muss nach seiner Entdeckung in genau einem Artikel veröffentlicht werden, aber über einen Artikel können beliebig viele Biomoleküle veröffentlicht werden. Ein Artikel hat einen eindeutigen Titel und hat ein Datum, an dem er veröffentlicht wurde. Artikel werden von mindestens einem oder mehreren Wissenschaftlern geschrieben. Ein Wissenschaftler hat einen eindeutigen Namen und kann beliebig viele Artikel schreiben. Ein Wissenschaftler hat Kontaktinformationen, die wiederum seine Telefonnummer und E-Mailadresse enthalten. Zu jedem DNA-Molekül wird die dazugehörige Nukleotidsequenz, die Strang-Orientierung und die Nummer des Chromosoms, auf dem sich das Molekül befindet, in der Datenbank gespeichert. Ein DNA-Molekül wird entweder zu keinem, einem oder verschiedenen mRNA-Molekülen transkribiert („übersetzt“). Der Bereich auf der DNA, der transkribiert wird, wird bei diesem Prozess durch eine Start- und Endposition beschrieben. Ein mRNA-Molekül gehört zu immer genau einem DNA-Molekül. Eine Nukleotidsequenz und ein sogenannter Vienna-String, der die Basenpaarung des mRNA-Moleküls repräsentiert, werden zu jedem mRNA-Molekül in der Datenbank erfasst. Ein mRNA-Molekül wird entweder zu keinem oder zu genau einem Protein synthetisiert, wobei ein Protein immer genau aus einem mRNA-Molekül entsteht. Ein Protein hat eine für das jeweilige mRNA-Molekül eindeutige Aminosäuresequenz und wird außerdem durch seine CATH-Klassifikation und sein Molekulargewicht in Kilodalton (kDa) beschrieben. Desweiteren können Proteine beliebig viele Domänen enthalten. Eine Domäne hat dabei eine eindeutige Domänen-ID und wird durch ein Hidden-Markov-Model (HMM) repräsentiert. Außerdem wird die Funktion jeder Domäne in der Datenbank hinterlegt. Eine Domäne kommt in mindestens einem Protein vor.*



	Lehrveranstaltung	<b>Grundlagen von Datenbanken</b>			WS 2013/14
	Aufgabenzettel	<b>3</b>			
	Gesamtpunktzahl	<b>40</b>			
	Ausgabe	<b>Mi. 13.11.2013</b>	Abgabe	<b>Do. 28.11.2013</b>	

## Aufgabe 3: Relationale Algebra und SQL

[9 P.]

Betrachten Sie das folgende relationale Datenbankschema.

<u>Rennfahrer</u>	<u>RID</u>	Vorname	Nachname	Geburt	Wohnort	<u>Rennstall</u>
	4	Sebastian	Vettel	1987-07-03	Kemmental (Schweiz)	2
	6	Fernando	Alonso	1981-07-29	Lugano (Schweiz)	5
	8	Marc	Webber	1976-08-27	Aston Clinton (UK)	2
	9	Lewis	Hamilton	1985-01-07	Genf (Schweiz)	31
	20	Jenson	Button	1980-01-19	Monte Carlo (Monaco)	31
	21	Felipe	Massa	1982-04-25	São Paulo (Brasilien)	5
	44	Kimi	Räikkönen	1979-10-17	Espoo (Finnland)	34


Rennstall → Rennstall.RSID

<u>Rennstall</u>	<u>RSID</u>	Name	Teamchef	Budget
	2	Red Bull	Christian Horner	370
	5	Ferrari	Stefano Domenicali	350
	31	McLaren	Martin Whitmarsh	220
	34	Lotus F1	Eric Boullier	100

<u>Rennort</u>	<u>OID</u>	Name	Strecke
	4	Australien GP	Albert Park Circuit
	15	Malaysia GP	Sepang International Circuit
	21	China GP	Shanghai International Circuit

<u>Platzierung</u>	<u>RID</u>	<u>OID</u>	Platz
	8	4	6
	4	15	1
	20	15	17
	4	4	3
	6	4	2
	8	15	2
	6	21	1
	9	4	5
	21	15	5
	20	4	9
	21	4	4

RID → Rennfahrer.RID, OID → Rennorte.OID

	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2013/14
	Aufgabenzettel	3			
	Gesamtpunktzahl	40			
	Ausgabe	Mi. 13.11.2013	Abgabe	Do. 28.11.2013	

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

- $\pi_{Nachname}(\sigma_{Platz="1"}((Platzierung) \bowtie Rennfahrer \bowtie \sigma_{Name="Malaysia GP"}(Rennort)))$
- $\pi_{Vorname, Nachname}(\sigma_{Budget < 350}(Rennstall) \bowtie_{RSID=Rennstall} Rennfahrer)$
- $\pi_{Rennstall.Name}(\pi_{RID}(\sigma_{Name="Australien GP"}(Rennort) \bowtie Platzierung) \bowtie_{Rennstall=RSID} Rennfahrer)$

- 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.]

- Namen der Rennställe, in denen junge Fahrer (Geboren in oder nach 1985) angestellt sind
- Persönliche Daten (Vor- und Nachname, Geburtsdatum) aller beim *Australien GP* platzierten Rennfahrer, die bei *McLaren* angestellt sind.
- Alle Informationen der Rennfahrer, die bisher keine Platzierung erreicht haben.
- Vor- und Nachnamen der Rennfahrerkollegen (selber Rennstall) von Button.

- c) Übersetzen Sie die folgenden umgangssprachlich formulierten Anfragen in einen zugehörigen SQL-Ausdruck. Testen Sie den SQL-Ausdruck auf der Übungsdatenbank (die entsprechenden Schemata sind für jede Gruppe angelegt). [2 P.]

- Persönliche Daten (Vor- und Nachname, Geburtsdatum) aller beim *Australien GP* platzierten Rennfahrer, die bei *McLaren* angestellt sind.
- Vor- und Nachnamen der Rennfahrerkollegen (selber Rennstall) von Button.

## Aufgabe 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 höchsten 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 Wohnort. Jeder 25. Rennfahrerangestellte ist ein Projektleiter. Die Namen von Rennenorten sind eindeutig.

- $\pi_{Wohnort}(\sigma_{Name="ChinaGP"}(\sigma_{Platz < 11}((Rennfahrer \bowtie Platzierung) \bowtie Rennort)))$
- $\pi_{Wohnort}(Rennfahrer \bowtie (\sigma_{Platz < 11} Platzierung) \bowtie (\sigma_{Name="ChinaGP"} Rennort))$