

Name:

Vorname:

Geburtsdatum:

Matrikelnummer:

Platz:

**Nachklausur**  
**Informatische Werkzeuge in den Geistes-**  
**und Sozialwissenschaften 2**

12. Oktober 2023

	To be used for grading, do not write here										
prob.	1.1	1.2	1.3	2.1	2.2	3.1	3.2	4.1	4.2	Sum	grade
total	10	4	5	4	10	9	4	8	6	60	
reached											

Die „Lösungen“ der Aufgaben in diesem Dokument sollen den Studierenden als Anfangspunkt für die Beantwortung der Aufgaben dienen. Trotz aller Bemühungen kann es zu Unvollständigkeiten oder sogar Fehlern in den „Lösungen“ kommen. Da die Korrektur und Benotung der gestellten Aufgaben niemals lediglich auf einen „Vergleich mit der Musterlösung“ hinausläuft, ist dies auch nicht so schlimm. In jedem Fall sollten die Studierenden die Lösungen nachvollziehen und im Prinzip mittels des Lehrstoffs selbst verifizieren können.

Sollten Sie „Lösungen“ finden, die Sie nicht verstehen oder sogar für fehlerhaft halten diskutieren Sie diese am besten mit den Tutor\*innen oder auf dem Kursforum und benachrichtigen Sie die Lehrenden; wir werden sie dann gegebenenfalls baldigst korrigieren.

# 1 Datenbanken

## Aufgabe 1.1 (Buchverwaltung)

10 Pkt

Gegeben ist folgender Datensatz, der den Bestand einer Bibliothek modelliert:

10 Min

BuchID	Titel	AusgeliehenVon	Kundennummer
123	Bürgerliches Gesetzbuch	Magdalena Musterfrau	00334
137	Die Physiker	-	-
055	Darm mit Charme	-	-
249	Die Verwandlung	Jan Modaal	75187
005	Faust	Jan Modaal	75187

Die BuchID identifiziert ein Buch eindeutig (Primary Key). Sie können davon ausgehen, dass es jedes Buch nur ein Mal in der Bibliothek gibt.

Sie wissen außerdem, dass für einen gleichen Eintrag in der Spalte **AusgeliehenVon** auch der Eintrag in der Spalte **Kundennummer** gleich sein wird.

Ein Bindestrich in den Spalten **AusgeliehenVon** und **Kundennummer** signalisiert, dass dieses Buch derzeit nicht ausgeliehen ist (NULL bzw. None).

### 1. (Tabelle erstellen, 4 Pkt.)

Geben Sie die SQL-Befehle zum Erstellen einer Tabelle unter dem Namen „Books“ mit der obigen Struktur an. Sie müssen keine Einträge hinzufügen. Vermeiden Sie dabei eventuelle Probleme, falls bereits eine Tabelle namens „Books“ existiert.

Lösung:

```
DROP TABLE IF EXISTS Books;
CREATE TABLE Books(
BuchID INTEGER PRIMARY KEY,
Titel TEXT,
AusgeliehenVon TEXT,
Kundennummer INTEGER);
```

### 2. (Daten anpassen, 3 Pkt.)

San Zhang (Kundennummer 67673) möchte das Buch *Die Physiker* ausleihen. Geben Sie den SQL-Befehl an, der die obige Tabelle entsprechend anpasst.

Lösung:

```
UPDATE Books
SET AusgeliehenVon = "San_Zhang", Kundennummer = "67673"
WHERE BookID = 137;
```

### 3. (Redundanz, 3 Pkt.)

In der obigen Tabelle werden einige Daten mehrfach gespeichert. Begründen Sie, warum das (gerade für größere Datensätze) suboptimal ist.

Lösung: In der Tabelle wird sowohl AusgeliehenVon als auch die Kundennummer festgehalten, obwohl bekannt ist, dass diese immer korrelieren. Diese Informationen könnten also (z.B. in eine zweite Tabelle "Kunden") ausgelagert werden.

Die doppelte Vorhaltung der Daten ist problematisch weil so leichter Fehler / Inkonsistenzen entstehen und mehr Speicherplatz verbraucht wird, als nötig.

### Aufgabe 1.2 (SQL Injection)

4 Pkt

Beschreiben Sie das Prinzip hinter einer "SQL injection attack" und erklären Sie, wie bereits beim Designprozess dagegen vorgegangen werden kann. 4 Min

**Lösung:** Viele Programme, die mit Datenbanken interagieren, brauchen Input direkt von den Benutzer\*innen des Programms. Bei einer SQL Attacke nutzt die angreifende Partei diesen Fakt, um direkt SQL-code auf der Datenbank auszuführen, was zu Datenverlust oder Datenlecks führen kann.

Dies kann verhindert werden, indem das Programm niemals direkt den Input benutzt, sondern immer erst überprüft, ob es legitimer Input ist (dieser Prozess wird "sanitising user input" genannt.).  
Siehe auch:



### Aufgabe 1.3 (Primär- und Fremdschlüssel)

5 Pkt

Erklären Sie kurz die Begriffe *Primärschlüssel* ("primary key") und *Fremdschlüssel* ("foreign key") im Kontext von Datenbanken. 5 Min

Welche Bedingungen müssen die Werte diese Schlüssel jeweils erfüllen?

**Lösung: Primärschlüssel:** Primärschlüssel sind Spalten einer Tabelle in einer Datenbank und werden zur eindeutigen Identifizierung eines Datensatzes verwendet. Der Wert eines Primärschlüssels darf nicht null sein und muss in einer Tabelle einmalig sein, da er jeden Datensatz eindeutig kennzeichnet.

**Fremdschlüssel:** Bei einem Fremdschlüssel handelt es sich um eine Spalte einer Tabelle, die auf einen Primärschlüssel einer anderen oder aber derselben Tabelle verweist.

Dabei darf der Wert des Fremdschlüssels nur Werte annehmen, die in der entsprechenden Tabelle auch tatsächlich vorkommen (referentielle Integrität).

## 2 Bild

### Aufgabe 2.1 (Vektorgrafiken)

4 Pkt

Vektorgrafiken speichern Bildinformationen nicht in Form eines Rasters von Pixeln, sondern als Liste von Primitiven (Kreise, Rechtecke, Polygone, etc). 4 Min

Erläutern Sie, wie diese Grafiken trotzdem auf Raster-Bildschirmen angezeigt werden können. Was bedeutet in diesem Kontext der Begriff *Anti-Aliasing*?

*Lösung:* Vektorgrafiken müssen vor der Anzeige auf einem Rasterbildschirm rasterisiert werden. Dabei evaluiert ein Programm für jede Form (Rechteck, Kreis, Polygon, ...) ob ein gegebener Pixel innerhalb der Form oder außerhalb der Form ist und weist dem Pixel gegebenenfalls die jeweilige Farbe zu.

Bei binären Entscheidungen (Pixel ist innerhalb der Form oder außerhalb, ja oder nein) entstehen dabei unschöne Treppenstufen bei diagonalen Kanten. Das liegt daran, dass Pixel keine Punkte sind, sondern eine gewisse Ausdehnung haben. Um die visuelle Qualität zu verbessern, versuchen Anti-Aliasing-Algorithmen zu entscheiden, wieviel Prozent eines Pixels eine Form überlappt. Die zugewiesene Form ist dann eine gewichtete Summe aus der Farbe der Form und des Hintergrundes. Dadurch werden Kanten weicher.

### Aufgabe 2.2 (Punkt in Polygon)

10 Pkt

Ihnen liegt eine Python-Liste mit Tupeln vor, die alle jeweils eine Form beschreiben wie im folgenden Beispiel:

10 Min

```
all_my_shapes = [
    ("square", (1.0, 1.0), 1.0),
    ("rectangle", [(2.0,2.0), (2.0,1.0), (1.0,1.0), (1.0,2.0)]),
    # ...
]
```

Bei Rechtecken werden alle vier Eckpunkte in einer Liste angegeben, bei Quadraten nur der untere linke Punkt und die Kantenlänge. Für die Zwecke dieser Aufgabe gibt es keine Formen außer Rechtecken und Quadraten und ein Punkt ist ein Tupel mit zwei `float`-Einträgen, die die X- und Y-Koordinaten angeben, z.B. (3.14, 1.59).

Schreiben Sie eine Python-Funktion `check_in_bounds(shape, point)`, die eine Form wie oben beschrieben und einen Punkt entgegen nimmt und überprüft, ob dieser Punkt innerhalb der oder auf den Grenzen der Form liegt und den entsprechenden Wahrheitswert (i.e. `True` oder `False`) zurück gibt.

*Lösung:* Hier ist eine mögliche Lösung:

```
def check_in_bounds(shape, point):
    """Checks if the given point falls within the given shape."""
    if shape[0] == "square":
        xmin = shape[1][0]
        xmax = shape[1][0] + shape[2]
        ymin = shape[1][1]
        ymax = shape[1][1] + shape[2]
    elif shape[0] == "rectangle":
        xmin = min(shape[1][0][0], shape[1][1][0], shape[1][2][0], shape[1][3][0])
        xmax = max(shape[1][0][0], shape[1][1][0], shape[1][2][0], shape[1][3][0])
        ymin = min(shape[1][0][1], shape[1][1][1], shape[1][2][1], shape[1][3][1])
        ymax = max(shape[1][0][1], shape[1][1][1], shape[1][2][1], shape[1][3][1])

    x_in_bounds = xmin <= point[0] and point[0] <= xmax
    y_in_bounds = ymin <= point[1] and point[1] <= ymax
    return x_in_bounds and y_in_bounds
```

### 3 Semantic Web & Kulturelles Erbe

#### Aufgabe 3.1 (XPath)

9 Pkt

Gegeben ist folgender Ausschnitt aus einer XML-Datei zur Katalogisierung einer Bibliothek:

9 Min

```
<library>
  <section id="fiction">
    <shelf>
      <book>
        <title>The Gargleblotter Chronicles</title>
        <author>Bilbo Wobblebottom</author>
      </book>
      <book>
        <title>The Quest for the Wobbly Widget</title>
        <author>Grizzlebop McSnortle</author>
      </book>
      ...
    </shelf>
    ...
  </section>
  <section id="nonfiction">
    <shelf>
      <book>
        <title>The Banana Peel Conspiracy</title>
        <author>Jellybean McSquirtle</author>
      </book>
      <book>
        <title>The Laughter Quake</title>
        <author>Funnybunny Fizzletwist</author>
      </book>
      ...
    </shelf>
    ...
  </section>
  ...
</library>
```

Geben Sie jeweils einen gültigen XPath-Ausdruck für die folgenden Beschreibungen an, der in der vollständigen Datei (mehr Einträge als hier gezeigt sind) alle gewünschten Knoten anwählt.

- Alle Buchtitel
- Den Titel des dritten Buches im zweiten Regal in der ersten Sektion
- Alle Namen von Autor\*Innen von Büchern in der nonfiction Sektion

---

*Lösung:* Hier sind mögliche Lösungen:

- /library/section/shelf/book/title
- /library/section[1]/shelf[2]/book[3]/title
- /library/section[@id="nonfiction"]/shelf/book/author

Insbesondere erwarten wir keine Selektion von text() oder dergleichen.

---

#### Aufgabe 3.2 (RDF Tripel)

4 Pkt

Im Kontext von semantischen Netzen und Wissensverarbeitung wird häufig von RDF-Tripeln oder *semantischen Tripeln* gesprochen. Ein Beispiel:

4 Min

("Michael\_Kohlhase", "hat\_Beruf", "Professor")

Erklären Sie, warum hier meist *ausschließlich* auf Tripel zurück gegriffen wird und nicht zum Beispiel auch Quadrupel, Quintupel oder Sextupel verwendet werden, um komplexere Zusammenhänge zu modellieren.

---

*Lösung:* Die drei Elemente eines RDF-Tripels entsprechen den Elementen einer simplen Satzstruktur: Subjekt, Prädikat und Objekt. Es wäre selbstverständlich möglich, auch längere Tupel zu verwenden. Allerdings würde dies zu Problemen bei der Modellierung und der Abfrage von Wissen führen, weil hier unnötige Komplexität eingeführt wird.

---

## 4 Geistiges Eigentum

### Aufgabe 4.1 (Beatrice CC-Lizenz)

8 Pkt

Sie erhalten folgende Nachricht von ihrer guten Freundin Beatrice Beispiel, welche kürzlich ihre Liebe für Musik entdeckt hat:

8 Min

Hey, du! Ich kann einfach nicht aufhören, vor Aufregung zu strahlen!

Also, du wirst es nicht glauben, aber mein neuestes Album, "Musikalische Märchenreise", ist gerade rausgekommen, und es ist so, sooo toll! Die Inspiration dafür kam, als ich auf einer Pizza mit extra Knoblauchsoße eingeschlafen bin und von musikalischen Feen besucht wurde. Ich meine, Knoblauch und Musik - das ist einfach die perfekte Kombi, oder? xD Und das einzige, was ich dafür versprechen musste, war, dass die Musik nicht benutzt wird, um Dinge zu verkaufen, da hatten die Feen was gegen.

Das Album ist wirklich wie eine wilde Achterbahnfahrt durch das Land der Noten. Wir haben sogar ein Lied, bei dem wir Glockenspiele mit einem Hula-Hoop-Reifen spielen. Ich sag dir, das war eine Herausforderung, aber der Sound ist einfach magisch!

Aber das Beste ist, dass ich mir wünsche, dass dieses Album in die Welt hinausgeht und die Menschen genauso glücklich macht, wie es mich gemacht hat. Deshalb haben wir uns für eine besondere Art der Freigabe entschieden. Etwas, das es jedem erlaubt, die Songs zu nutzen und zu teilen, ohne sich Sorgen um die Rechte zu machen (solange klar ist, dass die Musik von den Feen und mir kommt). Schließlich sollten wir alle gemeinsam in der Knoblauch- und Musikwolke schweben, oder nicht? :D

Also, auf bald, wenn du auf meiner musikalischen Märchenreise mitkommen möchtest!

Nennen Sie eine Creative Commons Lizenz, die nach dieser Beschreibung für Beatrices neue Album in Frage kommen würde und eine, die *nicht* in Frage kommen würde. Begründen Sie Ihre Antworten (kurz, 1-2 Sätze).

---

*Lösung:* Die offensichtliche Lizenz hier ist CC BY-NC da Beatrice (und die Feen) auf Namensnennung bestehen und kommerzielle Nutzung verhindern wollen.

Eine Lizenz, die nicht in Frage kommt wäre zum Beispiel CCO, die diese Einschränkungen nicht hat. Aber korrekte Antworten könnten auch CC BY-ND etc. sein, die nicht auf beide Bedingungen eingehen.

---

### Aufgabe 4.2 (CC Kombinationen)

6 Pkt

6 Min

Die meisten Lizenzen Creative Commons setzen sich aus Kombinationen der so genannten Module mit den Kürzeln BY, NC, ND und SA zusammen. Ein Beispiel dafür ist die Lizenz CC BY-NC.

Erklären Sie jeweils kurz (!) welche Bedeutung diese Module jeweils tragen. Nennen Sie außerdem eine Kombination der Module, die üblicherweise *nicht* für Lizenzen benutzt wird und erklären Sie, warum nicht.

---

*Lösung:* Die Bedeutungen der Module sind wie folgt:

- BY – Namensnennung  
Der Name des Urhebers muss genannt werden.
- NC – Nichtkommerzialität  
Das Werk darf nicht für kommerzielle Zwecke verwendet werden.
- BY – Keine Bearbeitung  
Das Werk darf nicht verändert werden.
- BY – Weitergabe unter gleichen Bedingungen  
Das Werk muss nach Veränderungen unter der gleichen Lizenz weitergegeben werden.

Daraus ergibt sich, dass z.B. die Kombination CC BY-ND-SA nicht (bedeutungsvoll) existieren kann, weil sie einerseits darauf besteht, dass nach Bearbeitung die gleiche Lizenz verwendet wird, allerdings auch Bearbeitung generell untersagt.

---