

# 1. Übungsblatt (WS 2020)

3.0 VU Datenmodellierung / 6.0 VU Datenbanksysteme

## Informationen zum Übungsblatt

### Allgemeines

In diesem Übungsteil entwerfen Sie eine kleine Datenbank mittels EER-Diagrammen, überführen ein EER-Diagramm in ein Relationenschema, und üben den Umgang mit der relationalen Algebra und dem Relationenkalkül.

Lösen Sie die Beispiele **eigenständig** (auch bei der Prüfung und vermutlich auch in der Praxis sind Sie auf sich alleine gestellt)! Wir weisen Sie darauf hin, dass sämtliche abgeschriebene Lösungen mit 0 Punkten beurteilt werden (sowohl das “Original” als auch die “Kopie”).

Geben Sie ein einziges PDF Dokument ab (max. 5MB). Erstellen Sie Ihr Abgabedokument computerunterstützt. Wir akzeptieren **keine PDF-Dateien mit handschriftlichen Inhalten**.

Das Übungsblatt enthält 8 Aufgaben, auf welche Sie insgesamt 15 Punkte erhalten können.

### Deadlines

bis **05.11. 12:00Uhr** Upload der Abgabe über TUWEL  
ab **18.11. 13:00Uhr** Korrektur und Feedback in TUWEL verfügbar

### Weitere Fragen – TUWEL Forum

Sie können darüber hinaus das TUWEL Forum verwenden, sollten Sie inhaltliche oder organisatorische Fragen haben. **Posten Sie auf keinen Fall Ihre (partielle) Lösungen im Forum!**

### Änderungen im Ablauf bzgl. COVID-19

Wegen der andauernden Sondersituation werden heuer keine Sprechstunden zum Übungsblatt stattfinden. Bitte wenden Sie sich stattdessen verstärkt an das TUWEL Forum wenn Sie Probleme damit haben die Angaben zu verstehen oder wenn Sie technische Schwierigkeiten haben.

Wenn möglich empfehlen wir Ihnen auch das Forum zur Diskussion mit Ihren Kommilitonen zu Nutzen. Ein gemeinsames Analysieren von Problemen hilft erfahrungsgemäß allen Beteiligten dabei den Stoff besser zu verstehen.

## Aufgaben: EER-Diagramme

<b>Aufgabe 1 (EER-Diagramm erstellen)</b>
---

[3 Punkte]
------------

Sie arbeiten an einem Startup, dass das Buchmarketing revolutionieren wird. Nachdem sie sich mehrere Millionen Euro an Funding gesichert haben beschliessen Sie das Projekt auch tatsächlich umzusetzen. Sie beginnning mit dem Entwurf der Datenbank auf Basis ihrer bisherigen Notizen.

Zeichnen Sie aufgrund der vorliegenden Informationen (siehe nächste Seite) ein EER-Diagramm. Verwenden Sie dabei die in der Vorlesung vorgestellte Schreibweise, sowie die (min,max)-Notation. Es sind keine NULL-Werte erlaubt, und Redundanzen sollen vermieden werden. Manchmal kann es notwendig sein, zusätzliche künstliche Schlüssel einzuführen.

Eine Unterstützung bei der Erstellung von EER-Diagrammen bietet das Tool dia (<http://wiki.gnome.org/Apps/Dia>, binaries unter <http://dia-installer.de>; Achtung: im Diagramm Editor ER auswählen!). Sie können das EER-Diagramm aber natürlich mit jeder beliebigen Software erstellen.

**Beschreibung des zu modellierenden Sachverhalts:**

Das zentrale Element der Datenbank ist das Erfassen von Schriftstücken. Für jedes Schriftstück wird ein Titel (TITEL) und eine Seitenanzahl (SEITEN) gespeichert, wobei der Titel für jedes Schriftstück eindeutig ist. Zusätzlich wird zu jedem Schriftstück gespeichert ob es sich speziell um ein Sachbuch, ein Roman oder Lyrik handelt. Ein Schriftstück muss allerdings nicht unbedingt in eine der drei Kategorien fallen. Bei einem Roman wird dabei zusätzlich das Genre (GENRE) gespeichert. Sachbücher können über Romane handeln, allerdings maximal über 3 verschiedene Romane. Es ist natürlich auch möglich dass ein Sachbuch gar nicht über einen Roman handelt. Bücher werden von Autorinnen geschrieben. Eine Autorin hat immer mindestens ein Schriftstück geschrieben und ein Schriftstück hat mindestens eine Autorin. Eine Autorin wird eindeutig durch eine Kombination an Namen (NAME) und Geburtsdatum (GEBDAT) identifiziert.

Zu einem Schriftstück können verschiedene Editionen herausgegeben werden. Die Editionen eines Schriftstücks können untereinander an ihrer Editionsnummer (ED-NR) unterschieden werden. Zwischen verschiedenen Schriftstücken sind Editionsnummern allerdings nicht zwingend eindeutig. Zusätzlich wird einer Edition ein Jahr (JAHR) zugeordnet. Eine Edition hat mindestens eine Auflage. Die Auflagen einer Edition können mittels Auflagennummer (ANR) unterschieden werden. Weiters wird zu jeder Auflage gespeichert wer den Druck übernimmt (DRUCKEREI). Weiters sind manche Editionen auch Sondereditionen, welche zusätzlich zu einer normalen Edition noch einen Anlass (ANLASS) haben.

Editionen werden von einem Verlag herausgegeben. Einem Verlag ist ein Name (NAME) und ein Budget (BUDGET) zugeordnet, wobei der Name den Verlag eindeutig identifiziert sind. In jedem Verlag gibt es Abteilungen welche innerhalb des Verlags durch eine eindeutige Kombination von Zuständigkeitsbereich (BEREICH) und der Stadt der Niederlassung (STADT) identifiziert. In verschiedenen Verlagen können allerdings die gleichen Bereich/Stadt Kombinationen vorkommen. Abteilungen ohne einen Verlag gibt es nicht.

Ein Verlag bewirbt an einem konkreten Datum (DATUM), über bis zu 2 Marketing Kanäle, eine oder mehrere Zielgruppen. Ein Marketing Kanal hat einen Namen (KNAME) – über den der Kanal eindeutig identifiziert wird – und zugehörige Kosten (KOSTEN). Weiters kann ein Marketing Kanal unter anderem ein Soziales Netzwerk oder eine Zeitung sein. Bei der Zeitung die Auflage (AUFLAGE). Bei Sozialen Netzwerken wird noch die konkrete Plattform (PLATFORM) erfasst. Für eine Zielgruppe werden eine Bezeichnung (BEZ) und eine Altersgruppe (AGRP) erfasst, wobei die Bezeichnung für jede Zielgruppe eindeutig ist.

Zielgruppen nutzen Soziale Netzwerke und interessieren sich für mindestens ein Schriftstück. Weiters liebt jede Zielgruppe bis zu 5 Autorinnen. Leider hasst jede Zielgruppe auch mindestens eine (und beliebig viele) Autorinnen. Das bewegt Autorinnen dazu unter Pseudonymen zu arbeiten. Dazu wird erfasst welche Autorin ein Alias von welcher anderen Autorin ist.

**Aufgabe 2 (Semantik von EER Diagrammen)****[1 Punkt]**

Betrachten Sie das in Abbildung 1 dargestellte EER-Diagramm.

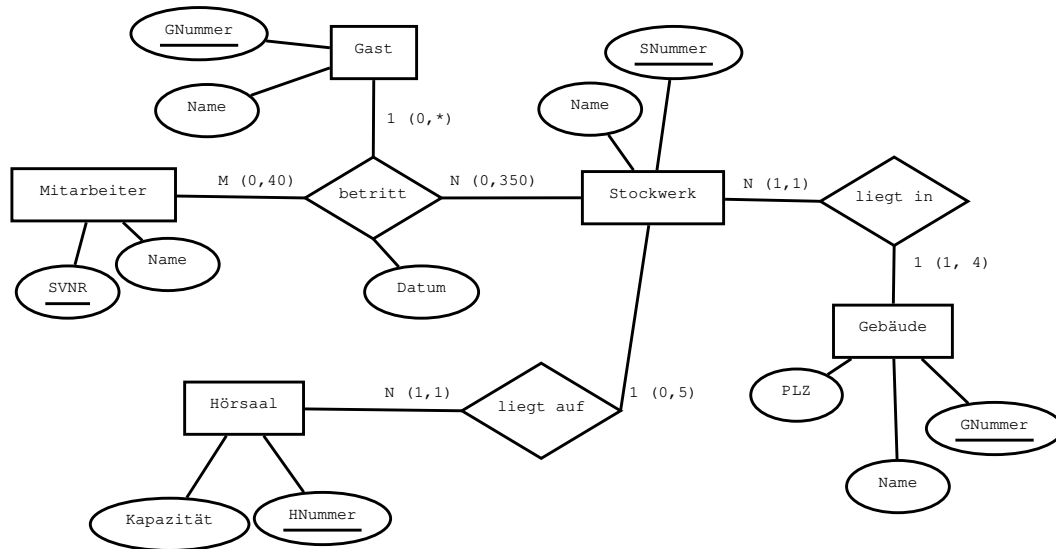


Abbildung 1: EER-Diagramm zu Aufgabe 2

In dem ER-Diagramm wird sowohl die Notation mittels Funktionalitäten, als auch die (min,max)-Notation verwendet.

(Anmerkung: dies geschieht hier zu Übungszwecken und ist in der Praxis nicht üblich.)

Das Diagramm enthält dadurch mehr Informationen als wenn nur eine der beiden Notationen verwendet worden wäre.

- Geben Sie einen konkreten Beziehungstyp im Diagramm an, bei welchem das Weglassen einer Notation zum Verlust von Informationen führt.
- Für den gewählten Beziehungstyp führt das Weglassen welcher Notation zum Informationsverlust?
- Erklären Sie kurz in eigenen Worten, welche Information nicht mehr dargestellt werden kann.
- Geben Sie ein konkretes Beispiel für die verlorene Information an. D.h., für den von Ihnen gewählten Beziehungstyp, geben Sie eine Ausprägung an welche (mindestens) eine durch die weggelassene Notation ausgedrückte Bedingung verletzt, aber sämtliche Beschränkungen der verbleibenden Notation erfüllt.

**Aufgabe 3 (Überführung ins Relationenschema)****[2 Punkte]**

Überführen Sie das EER-Diagramm aus Abbildung 2 in ein Relationenschema. Nullwerte sind nicht erlaubt (Sie können dabei annehmen, dass alle für einen Entitätstyp angegebenen Attribute für alle Entitäten dieses Typs existieren; d.h. die Definiertheit sämtlicher Attribute ist 100%). Verwenden Sie möglichst wenig Relationen. Unterstreichen Sie sämtliche Primärschlüssel, schreiben Sie die Fremdschlüssel kursiv und stellen Sie sicher, dass ein Fremdschlüssel eindeutig der passenden Relation zugeordnet werden kann.

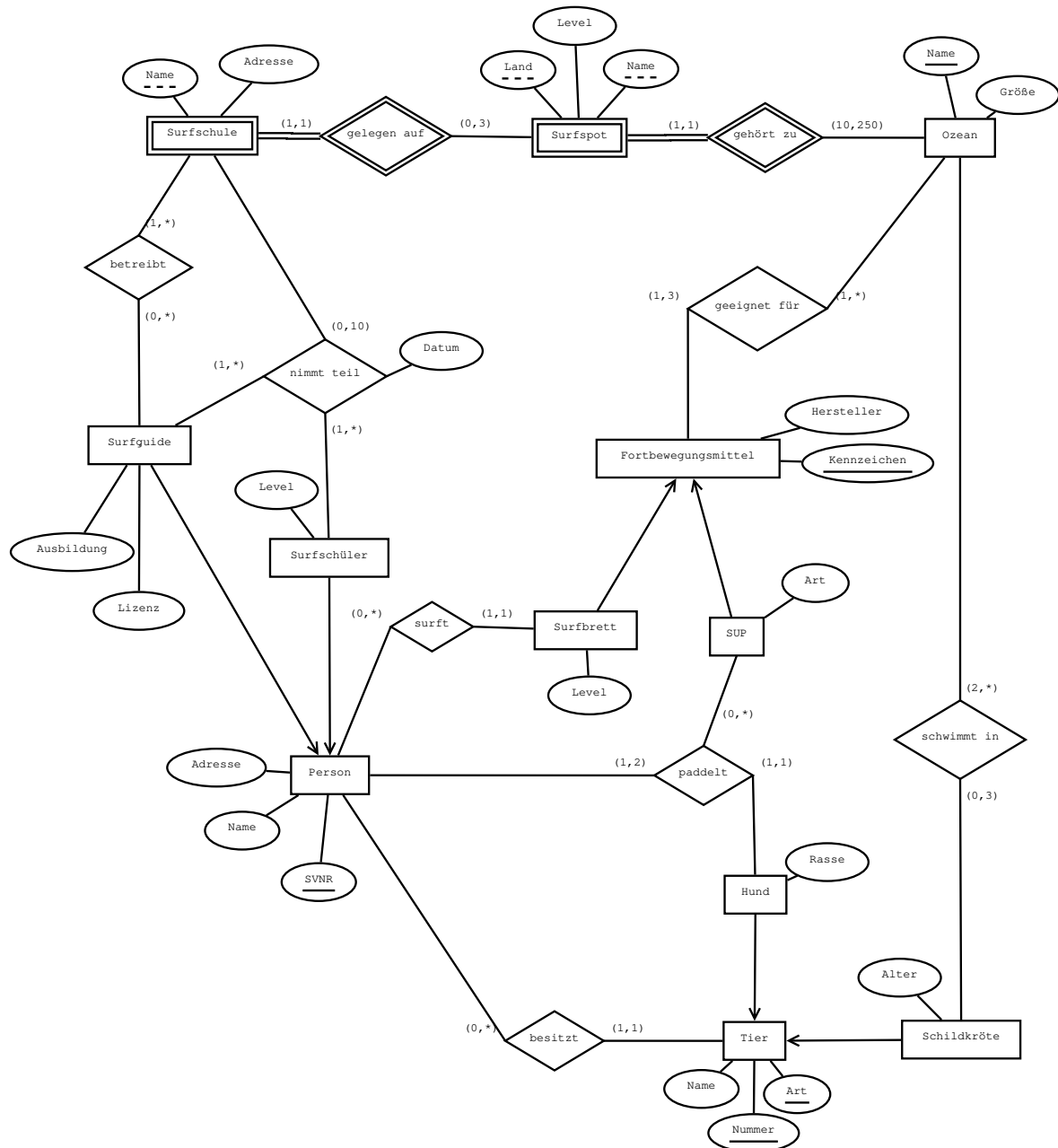


Abbildung 2: EER-Diagramm zu Aufgabe 3

## Aufgaben: Relationale Algebra - Relationenkalkül

Um Ihnen die Erstellung Ihrer Abgabe zu den folgenden Aufgaben zu erleichtern, haben wir unter <http://dbai.tuwien.ac.at/education/dm/resources/symbols.html> eine Liste mit den wichtigsten Symbolen der relationalen Algebra zusammengestellt. Sie können diese per copy/paste in Ihr Word/LibreOffice/OpenOffice/...Dokument einfügen. Zusätzlich sind die entsprechenden L<sup>A</sup>T<sub>E</sub>X Befehle vermerkt.

### Aufgabe 4 (Auswerten)

[0.5 Punkte]

Gegeben sind die folgenden vier Relationen.

Band		
name	genre	founded
Cid Rim	TBD	2010
JSBL	Funk	2008
CSH	Indie	2010
Dorian Concept	Electronic	2006

Musician		
name	instrument	studio
Will	Guitar	False
Peter	Bass	True
Clemens	Drums	True
Dorian	Keyboard	True

partOf	
mname	bname
Clemens	JSBL
Clemens	Cid Rim
Will	CSH
Dorian	JSBL
Dorian	Dorian Concept

Song			
title	by	length	genre
Killer Whales	CSH	6:14	Indie
My Boy	CSH	2:52	Indie
Swnerve	Cid Rim	4:26	Electronic
Failures III	Dorian Concept	3:40	Jazz

Bestimmen Sie das Ergebnis der folgenden Anfragen über diesen Relationen.

(a)

$$\pi_{\text{title, genre}} \left( \pi_{\text{title, length}}(\text{Song}) \times ((\rho_{\text{mname} \leftarrow \text{name}}(\sigma_{\text{studio}=\text{True}}(\text{Musician})) \bowtie \text{partOf}) \bowtie \rho_{\text{bname} \leftarrow \text{name}}(\sigma_{\text{founded} \geq 2010}(\text{Band}))) \right)$$

(b)

$$\{m.\text{name}, b.\text{name} \mid m \in \text{Musician} \wedge b \in \text{Band} \wedge m.\text{instrument} \neq \text{'Bass'} \wedge \exists p \in \text{partOf} (b.\text{name} = p.\text{bname} \wedge m.\text{name} = p.\text{mname} \wedge (\forall s \in \text{Songs} (b.\text{name} = s.\text{by} \rightarrow s.\text{genre} = b.\text{genre})))\}$$

**Aufgabe 5 (Äquivalenzen)**

[2 Punkte]

Gegeben sind die folgenden Relationenschemata  $R(\underline{ABC})$ ,  $S(\underline{BDE})$ ,  $T(\underline{CDF})$  sowie untenstehende Paare  $q_i, q_j$  an Ausdrücken der relationalen Algebra. Für alle diese Paare:

- Überprüfen Sie, ob die jeweiligen Ausdrücke äquivalent sind (also ob Sie über allen möglichen Ausprägungen der Schemata immer das gleiche Ergebnis liefern). Sie können dabei davon ausgehen, dass NULL-Werte in den Ausprägungen verboten sind.
  - Begründen Sie Ihre Antwort mit einer kurzen **Erklärung**.
  - Falls die beiden Ausdrücke *nicht* äquivalent sind, geben Sie zusätzlich noch ein **Gegenbeispiel** an. (Ein Gegenbeispiel besteht aus konkreten Ausprägungen der beteiligten Relationenschemata sowie den Ergebnissen beider Ausdrücke über diesen Ausprägungen.) Das Gegenbeispiel kann entfallen wenn einer der beiden Ausdrücke kein gültiger Ausdruck der Relationalen Algebra ist. In diesem Fall reicht die Erklärung aus.
- (a)  $q_1: (\pi_C(T) - \sigma_{C>5}(R)) \cap \rho_{C \leftarrow B}(\pi_B(S))$  und  
 $q_2: (\sigma_{C>5}(\pi_C(T) - \pi_C(R))) \cap \rho_{C \leftarrow B}(\pi_B(S))$
- (b)  $q_3: S \bowtie (\pi_{CD}(T) \cap \rho_{D \leftarrow A}(\pi_{CA}(R)))$  und  
 $q_4: (S \bowtie \pi_{CD}(T)) \cap (S \bowtie \rho_{D \leftarrow A}(\pi_{CA}(R)))$
- (c)  $q_5: \pi_{AB}(\sigma_{A>B \vee A<B}(R)) \times \pi_{DF}(\sigma_{D>F \vee D<F}(T))$  und  
 $q_6: \pi_{ABDF}(\sigma_{(A>B \wedge D>F) \vee (A<B \wedge D<F)}(R \times T))$

**Aufgabe 6 (Größenabschätzung)**

[1.5 Punkte]

Gegeben sind die Relationenschemata  $R(\underline{AB})$ ,  $S(\underline{ABCD})$ , und  $T(\underline{ACE})$  sowie je eine Ausprägung für jedes dieser Schemata, wobei  $|R|$  Tupel in der Ausprägung für  $R$  vorhanden sind,  $|S|$  Tupel in jener für  $S$ , und  $|T|$  Tupel in jener für  $T$ .

- Geben Sie die minimale bzw. maximale Größe (= Anzahl der Tupel) folgender Ausdrücke unter Annahme der angegebenen Werte für  $|R|$ ,  $|S|$ ,  $|T|$  in relationaler Algebra an.
  - Begründen Sie Ihre Antwort.
  - Geben Sie sowohl für das Minimum als auch für das Maximum jeweils **konkrete Ausprägungen** mit der angegebenen Anzahl von Tupeln an ( $|R|$  für  $R$ ,  $|S|$  für  $S$ ,  $|T|$  für  $T$ ), auf welchen die Anfrage tatsächlich die von Ihnen berechnete minimale/maximal Anzahl an Tupeln enthält. Geben Sie jeweils außerdem das Ergebnis der Anfrage für Ihre Ausprägungen an.
- (a)  $q_1: \rho_{B \leftarrow C}(T) \bowtie \sigma_{A=4 \wedge B=2}(R)$  (mit  $|R| = 5$  und  $|T| = 4$ )
- (b)  $q_2: \pi_{EX}(\rho_{X \leftarrow A, Y \leftarrow C}(S) \times T) - \pi_{EX}(\rho_{X \leftarrow A}(T) \bowtie S)$  (mit  $|S| = 7$  und  $|T| = 4$ )
- (c)  $q_3: (\pi_A(\sigma_{A \neq 1}(R)) \cap \pi_A(\sigma_{A>3}(T))) \cup \rho_{A \leftarrow C}(\pi_C(T) \bowtie \pi_C(S))$   
(mit  $|R| = 5$ ,  $|S| = 3$  und  $|T| = 3$ )

**Aufgabe 7 (Abfragesprachen)**

[1 Punkt]

Gegeben sind die Relationenschemata  $R(\underline{AB})$ ,  $S(\underline{ABCD})$ , und  $T(\underline{ACE})$ .

In den folgenden Aufgaben ist eine Abfrage in einer der Sprachen aus der Vorlesung gegeben. Die Abfrage soll in die jeweils beiden anderen Sprachen übersetzt werden.

- (a) Übersetzen Sie die Abfrage

$$\pi_{A,C}(S \bowtie T)$$

in den Tupel- und den Domänenkalkül.

- (b) Übersetzen Sie die Abfrage

$$\{x \mid \exists y, z ([x, y] \in R \wedge [y, x, z, 3] \in S)\}$$

in den Tupelkalkül und in Relationale Algebra.

- (c) Übersetzen Sie die Abfrage

$$\{u, v \mid u \in R \wedge v \in T \wedge \exists w \in S (w.C > v.C)\}$$

in den Domänenkalkül und Relationale Algebra.

**Aufgabe 8 (Formalisieren von Anfragen)**

[4 Punkte]

Ein Unternehmen verwaltet Informationen über die Zutritte seiner Mitarbeiter in einer Datenbank mit folgendem Schema (Primärschlüssel sind unterstrichen, Fremdschlüssel sind kursiv geschrieben).

Mitarbeiter (PersNr, SVNr, MName)

Stockwerk (SName, Bezeichnung)

betritt (MitarbeiterNr: *Mitarbeiter.PersNr*, StockName: *Stockwerk.SName*, Datum)

verlässt (MitarbeiterNr: *Mitarbeiter.PersNr*, StockName: *Stockwerk.SName*, Datum)

Gebäude (GName, PLZ)

unterteiltIn(StockwerkN: *Stockwerk.Name*, GName: *Gebäude.GName*)

Hörsaal (HNummer, Name)

liegtIn (HörsaalNr: *Hörsaal.HNummer*, SName: *Stockwerk.SName*)

(Sie dürfen im Folgenden gerne passende (eindeutige) Abkürzungen sowohl für die Relationen als auch die Tabellennamen verwenden.)

Formulieren Sie die unten beschriebenen Abfragen jeweils sowohl in der **relationaler Algebra**, dem **Tupelkalkül** und dem **Domänenkalkül**.

- (a) Geben Sie den Namen, die Personalnummer und die Sozialversicherungsnummer aller Mitarbeiter an, die sich jetzt gerade auf Stockwerk mit Namen 'H1' befinden.
- (b) Es befinden sich gerade Studenten in Hörsaal mit dem Name 'Gödel'. Geben Sie die Mitarbeiter aus, die gerade auf einem Stockwerk sind, auf welchem sich auch diese Studenten aufhalten.
- (c) Geben Sie alle Mitarbeiter aus, die an einem Tag in verschiedenen Gebäuden waren.