

**Bergische Universität Wuppertal**  
**Schumpeter School of Business and Economics**

**Bachelor of Science / Bachelor of Arts**

**WS 2020/21**

---

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien BWiGes 5.8 Operations Management und Informationstechnologien Grundlagen der Wirtschaftsinformatik Modul I (PO Neufassung 2014)
Tag der Prüfung:	17.02.2021
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar), Aufgabenblock A: beigefügte Formelsammlung, Aufgabenblock D: Anwendungsbeispiele 1 und 2.

---

**Bearbeiten Sie 2 der angegebenen 4 Aufgabenblöcke vollständig!**

Die Lösungen zu den Aufgaben sollen gegliedert und in vollständigen zusammenhängenden Sätzen dargestellt und Rechnungen mit ihren Zwischenschritten nachvollziehbar sein. Die Darstellungsform und die Systematik der Gedankenführung gehen in die Bewertung ebenfalls ein. In Klammern ist für jede Aufgabe die Anzahl der maximal möglichen Punkte angegeben, die bei einer richtigen und vollständigen Bearbeitung erreicht werden können. Sie entspricht in etwa dem erwarteten Zeitbedarf in Minuten. Insgesamt können **90 Punkte** erreicht werden. Für eine erfolgreiche Bearbeitung müssen wenigstens **45 Punkte** erworben werden.

**Für Bachelor-Studiengänge ist der Aufgabenblock A verbindlich!**

Bei Bearbeitung von mehr als zwei Aufgabenblöcken wird die Bearbeitung des Aufgabenblocks A und des ersten weiteren bearbeiteten Blocks gewertet.

Die Klausur besteht mit diesem Deckblatt aus insgesamt 10 (zehn) Seiten.

# Aufgabenblock A:

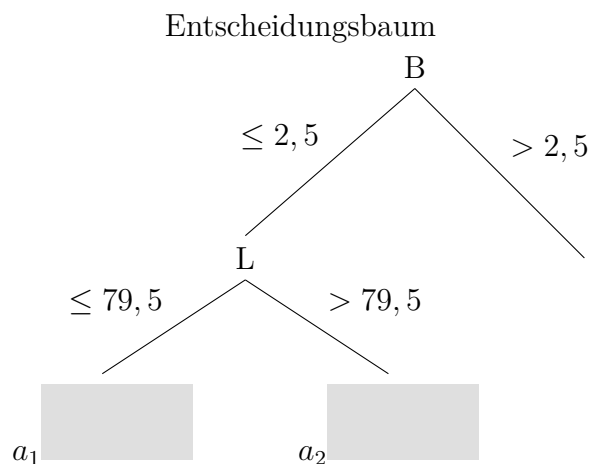
## Knowledge-based systems

### Aufgabe A.1: Entscheidungsbäume

[15 Punkte]

Gegeben ist ein Datensatz mit PKW-Daten mit den drei Attributen **E**rstzulassungsjahr, **L**aufleistung in tausend Kilometer und Anzahl der **B**esitzer sowie die Klassifikation in die **R**isikoklassen **G**ering, **M**ittel und **H**och. Zusätzlich wurde mittels der CART Methodik bereits ein zugehöriger, unvollständiger Entscheidungsbaum erstellt.

Datensatz				
i	E	L	B	R
1	2016	39	1	G
2	2015	41	2	G
3	2016	50	3	M
4	2017	29	4	H
5	2005	54	3	M
6	2007	79	2	G
7	2013	80	1	H
8	2015	62	4	M
9	2010	15	2	G



- a) Vervollständigen Sie die linke Seite des Entscheidungsbaums, indem Sie für die beiden Blattknoten  $a_1$  und  $a_2$  die resultierende Klassifikation in den grauen Kästchen notieren. (2 Punkte)
- b) Bestimmen Sie mittels des *Gini impurity*-Maßes den Test, der zur nächsten Verzweigung der rechten Baumhälfte am geeignetsten ist. Testen Sie jedes Attribut (E, L, B) nur einmal für die Verzweigung, mittels der folgenden Tests:
- $E \leq 2015,5 / E > 2015,5$
  - $B \leq 3,5 / B > 3,5$
  - $L \leq 39,5 / L > 39,5$

Erweitern Sie anschließend den obigen Entscheidungsbaum. Klassifiziert der nun erweiterte Entscheidungsbaum den Datensatz vollständig und fehlerfrei? Diskutieren Sie abschließend kurz die Qualität des Entscheidungsbaumes hinsichtlich der Problematik *Overfitting* vor dem Hintergrund, dass der Entscheidungsbaum in Zukunft alle 10.000 PKWs einer Ortschaft klassifizieren soll. (9 Punkte)

- c) Geben Sie einen weiteren Test bezogen auf eins der Attribute E, L oder B (abseits der in Teilaufgabe b) dargestellten) an, der die Fälle mit  $B > 2,5$  eindeutig klassifiziert. (4 Punkte)

**Aufgabe A.2: Thesen**

[15 Punkte]

Nehmen Sie zu den folgenden Thesen begründet Stellung. Eine auf „ja“ oder „nein“ beschränkte Antwort enthält keine Punkte.

- a) Wir betrachten die  $t$ -te Iteration des Verfahrens *Adaboost M.2* für einen Datensatz mit  $C = 4$  Klassenlabels. In dieser Iteration wird der schwache Schätzer  $h_t$  mit minimalem gewichteten Fehler  $\epsilon_t = 0,6$  ausgewählt. Da in diesem Fall der schwache Schätzer  $h_t$  schlechtere Ergebnisse liefert als einfaches Raten, terminiert das Verfahren und gibt die Schätzfunktion  $h_{fin}(x)$ , unter Verwendung der Schätzer  $h_1, \dots, h_{t-1}$  aus. (5 Punkte)
- b) Wir betrachten das Clustering mittels des Algorithmus' DBSCAN und folgendes Szenario: Während des Verfahrens wird der Datensatz  $x_i$  als *Noise* markiert, da dieser keine ausreichend große  $\epsilon$ -Nachbarschaft, abhängig vom Parameter  $p^{min}$  besitzt. Dies bedeutet, dass der Punkt  $x_i$  auch am Ende des Verfahrens als *Noise* zählt und keinem Cluster zugeordnet ist. (5 Punkte)
- c) Wir betrachten die beiden Ensemble-Methoden Boosting und Bagging. Beide Verfahren kombinieren mehrere Schätzer zu einer Gesamtschätzung. Dabei gilt für beide Methoden, dass die einzelnen Schätzer gleich gewichtet in die Gesamtschätzung einfließen. (5 Punkte)

**Aufgabe A.3: Clustering**

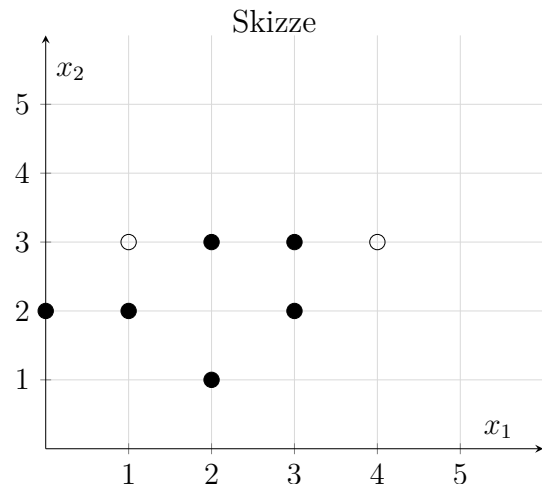
[15 Punkte]

Im Folgenden wird das *k-means*-Clusteringverfahren betrachtet.

Gegeben sind Datenpunkte  $x_1, \dots, x_6$  (skizziert als ●) sowie die beiden Zentren  $C = \{c_1, c_2\}$  (skizziert als ○).

Daten $x_i$		
i	$x_1$	$x_2$
1	0	2
2	1	2
3	2	1
4	3	2
5	3	3
6	2	3

Zentren $c_i$		
i	$x_1$	$x_2$
1	1	3
2	4	3



Hinweis: Verwenden Sie zur Bearbeitung der Teilaufgaben das *Manhattan-Distanzmaß*.

- a) Bestimmen Sie die aktuellen Cluster  $C_1$  und  $C_2$ . (2 Punkte)
- b) Führen Sie eine weitere Iteration des *k-means*-Verfahrens durch und geben sie abschließend die neuen Cluster  $C_1$  und  $C_2$  sowie die neuen Zentren  $c_1$  und  $c_2$  an. Skizzieren Sie die beiden Zentren zusätzlich in obiger Skizze mit dem Symbol  $\otimes$ . (9 Punkte)
- c) Zeigen Sie anhand einer selbst gewählten Menge von Punkten aus obigem Datensatz  $(x_1, \dots, x_6)$ , dass die Minkowski-Distanz für  $\alpha = 0,75$  keine Metrik ist. (4 Punkte)

## Formelsammlung zum Aufgabenblock A

$$\text{gain}(M, A_j) = \max \{ \text{gain}(M, A_l) \mid l \in \{1, \dots, k\} \}$$

$$\text{gain}(M, A_j) = \text{info}(M) - \text{info}(M, A_j)$$

$$\text{info}(M) = - \sum_{i=1}^k \frac{\text{freq}(C_i, M)}{\text{freq}(M)} \cdot \log_2 \left( \frac{\text{freq}(C_i, M)}{\text{freq}(M)} \right)$$

$$\text{info}(M, A_j) = \sum_{i=1}^l \frac{\text{freq}(T_i)}{\text{freq}(M)} \cdot \text{info}(T_i)$$

$$\text{prob}(C_i, M) = \frac{\text{freq}(C_i, M)}{\text{freq}(M)} = \frac{\text{freq}(C_i, M)}{|M|}$$

$$G(b_l) = \frac{\#(b_l, l)}{m} \cdot \left( 1 - \sum_{i=1}^k N(i, b_l, l)^2 \right) + \frac{\#(b_l, r)}{m} \cdot \left( 1 - \sum_{i=1}^k N(i, b_l, r)^2 \right)$$

$$\forall d^\alpha(\mathbb{X}_1, \mathbb{X}_2) = \left( \sum_{j=1}^d |x_{1,j} - x_{2,j}|^\alpha \right)^{1/\alpha}$$

$$w_{i,j} = \frac{1}{\sum_{k=1}^n \left( \frac{\|x_j - \mu_i\|}{\|x_j - \mu_k\|} \right)^{\frac{2}{m-1}}} = \frac{1}{\|x_j - \mu_i\|^{\frac{2}{m-1}} \cdot \sum_{k=1}^n \|x_j - \mu_k\|^{\frac{-2}{m-1}}}$$

$$\mu_i = \sum_{j=1}^n \frac{(w_{i,j})^m}{\sum_{j=1}^n (w_{i,j})^m} \cdot x_j = \frac{1}{\sum_{j=1}^n (w_{i,j})^m} \cdot \sum_{j=1}^n (w_{i,j})^m \cdot x_j$$

$$c_i = \frac{1}{|C_i|} \cdot \sum_{x \in C_i} x$$

$$p(x') = \frac{D(x')^2}{\sum_{x \in X} D(x)^2}$$

## Aufgabenblock B:

### Computer Hardware und Systembetrieb

#### (Grundlagen der Rechnerarchitektur und Informationsverarbeitung)

##### Aufgabe B.1: Pipeline

[5 Punkte]

Erläutern Sie die Begriffe Stalling, Bubble und (Pipeline-)Flush.

##### Aufgabe B.2: Speculative Execution

[10 Punkte]

- a) Was versteht man unter Speculative Execution? (3 Punkte)
- b) Welche Aktionen sind erforderlich, wenn Befehle aufgrund falscher Sprungvorhersagen spekulativ fälschlich ausgeführt wurden? (3 Punkte)
- c) Welche Arten von Operationen sind dabei besonders problematisch? (4 Punkte)

##### Aufgabe B.3: SMT vs. Multi-Core

[10 Punkte]

Erläutern Sie die Hauptunterschiede zwischen einem Prozessor mit 2 Kernen und einem Prozessor mit einem Kern und 2-Weg SMT.

##### Aufgabe B.4: Copy-on-Write

[10 Punkte]

- a) Erläutern Sie kurz die Idee hinter Copy-on-Write. (6 Punkte)
- b) Wo wird dieses Konzept typischerweise eingesetzt? (4 Punkte)

##### Aufgabe B.5: HPC-Cluster/COTS-Produkte

[10 Punkte]

- a) Was versteht man unter COTS-Produkten? (4 Punkte)
- b) Wieso verwendet man heute oft Standardtechnologien anstelle spezieller Entwicklungen für den HPC-Bereich? (3 Punkte)
- c) Für welche Komponenten eines HPC-Clusters lassen sich COTS-Produkte nutzen? (3 Punkte)

# Aufgabenblock C:

## Kommunikationssysteme (Internet-Technologien)

### Aufgabe C.1: Internet [15 Punkte]

- a) Erläutern Sie die Aufgaben und Eigenschaften der einzelnen Schichten des Internet-Referenzmodells. (5 Punkte)
- b) Nennen und erläutern Sie mindestens fünf Vorteile von IPv6 gegenüber IPv4. (10 Punkte)

### Aufgabe C.2: World Wide Web [15 Punkte]

- a) Erläutern Sie kurz folgende Standards und Technologien im WWW:
- XHTML
  - SOAP
  - RIA
  - CSS
  - DOM (5 Punkte)
- b) Stellen Sie in einer Skizze die HTTP-Interaktion zwischen einem Webbrowser und einem Webserver mit allen beteiligten Funktionseinheiten dar, wobei die angeforderte DHTML-Ressource dynamisch per Java-Servlet generiert wird. Beschreiben Sie kurz alle Arbeitsschritte von der Ressourcenanforderung bis zum Abschluss des Seitenaufbaus (Rendering). (6 Punkte)
- c) Erstellen Sie eine vergleichende Skizze der jeweils genutzten Protokollstapel im WWW
- Bei Nutzung regulärer unverschlüsselter HTTP-Kommunikation,
  - bei Nutzung eines Webservice auf SOAP-Basis,
  - bei HTTP-Kommunikation in einem Overlay-Network via SSL-VPN mit in HTTPS gekapselten IP-Paketen und
  - bei HTTP-Kommunikation in einem VPN auf IPsec-Basis mit Tunnelmodus. (4 Punkte)

### Aufgabe C.3: E-Mail [15 Punkte]

- a) Skizzieren und beschreiben Sie das Speichervermittlungsverfahren für die Übermittlung einer E-Mail vom E-Mail-Client des Absenders zum E-Mail-Client des Empfängers. (5 Punkte)
- b) Um welche Funktionalität werden E-Mails durch den MIME-Standard erweitert? (2 Punkte)
- c) Beschreiben Sie, wie sich E-Mail-Anhänge für sichere und vertrauenswürdige E-Mails nutzen lassen. Welche Daten werden in einer per S/MIME hybrid verschlüsselten und unterzeichneten E-Mail übertragen und welche Schlüssel sind an welcher Stelle erforderlich? (8 Punkte)

## Aufgabenblock D:

### Datenorganisation (Datenbankmanagementsysteme)

Bearbeiten Sie 3 der folgenden 4 Aufgaben! (insgesamt 45 Punkte)

**Aufgabe D.1: Architektur** [15 Punkte]

- a) Erläutern Sie die strukturellen Grundelemente des Relationenmodells (Relation, Tupel, Attribut, Datentyp) in ihrem Gesamtzusammenhang. Verwenden Sie für Ihre Erläuterungen eine selbst gewählte Beispielrelation. (10 Punkte)
- b) Skizzieren Sie das Grundgerüst eines Klassifikationsschemas für Integritätsbedingungen im Relationenmodell. (5 Punkte)

**Aufgabe D.2: Interne Ebene** [15 Punkte]

- a) Erläutern Sie die Grundfunktion der zeilenorientierten Speicherung in Relationalen Datenbanksystemen. Verwenden Sie für Ihre Argumentation die folgenden Begriffe: stored record, stored file, page, page set. (6 Punkte)
- b) Erläutern Sie das Konzept des Mehrebenenindex (Multilevel-Index) am Beispiel des  $B^*$ -Baums. Stellen dabei auch heraus, welche Teile des  $B^*$ -Baums einen *dense index* respektive einen *non-dense index* repräsentieren. (9 Punkte)

**Aufgabe D.3: Sicherheit** [15 Punkte]

- a) Erläutern Sie die Grundkonzepte von Authentifizierung und Autorisierung. (9 Punkte)
- b) Erläutern Sie das Konzept des *Label Based Access Control*. (6 Punkte)

*Hinweis: Aufgabe D.4 folgt auf der nächsten Seite ...*

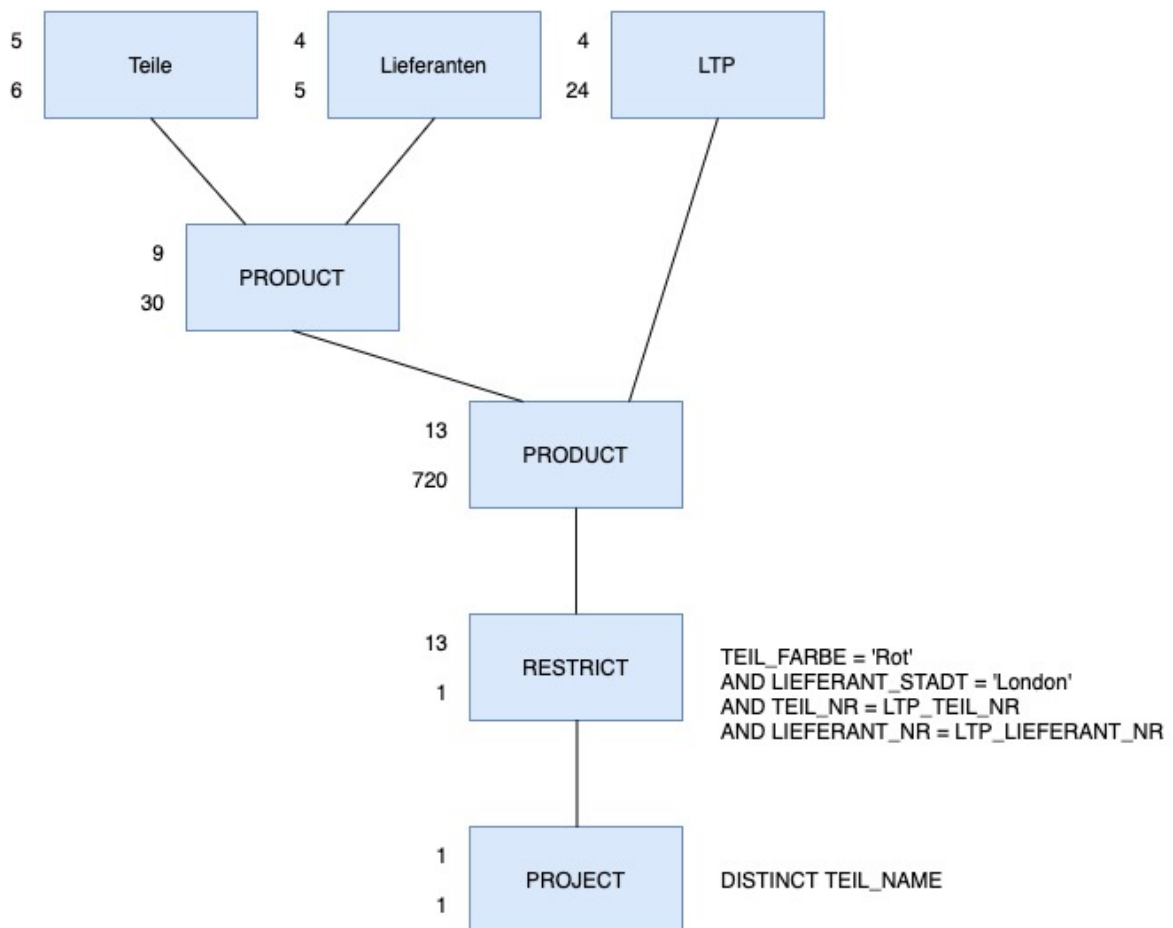
### Aufgabe D.4: Anfrageoptimierung

[15 Punkte]

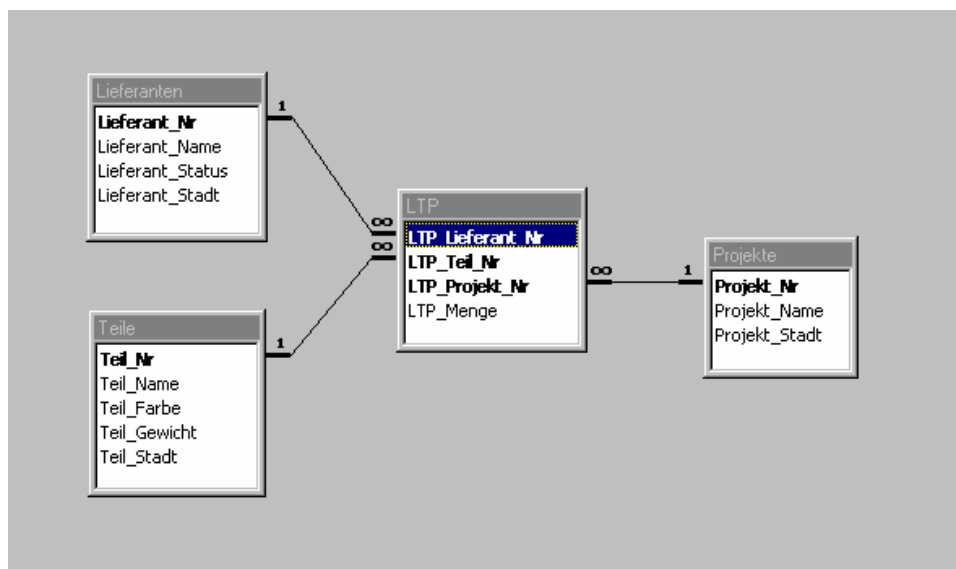
Die unten stehende Abbildung zeigt einen initialen logischen Anfrageplan (iLAP) (mit Angaben zum Mengengerüst mit Grad (oberer Wert) und Kardinalität (unterer Wert) für jede Ausgangs- oder Ergebnisrelation), der auf dem angegebenen SQL-Ausdruck basiert.

Entwickeln Sie aus dem iLAP einen präferierten logischen Anfrageplan (pLAP). Geben Sie für jeden Knoten des pLAP - sofern möglich - Grad und Kardinalität an. Benennen Sie darüber hinaus die von Ihnen verwendeten Gesetzmäßigkeiten und Heuristiken der Anfrageoptimierung in Relationalen Datenbanksystemen, mit deren Hilfe Sie den pLAP entwickelt haben!

```
SELECT DISTINCT TEIL_NAME
FROM TEILE, LIEFERANTEN, LTP
WHERE TEIL_FARBE = 'Rot'
AND LIEFERANT_STADT = 'London'
AND TEIL_NR = LTP_TEIL_NR
AND LIEFERANT_NR = LTP_LIEFERANT_NR
```







## Lieferanten

Lieferant_Nr	Lieferant_Name	Lieferant_Status	Lieferant_Stadt
L1	Smith	20	London
L2	Jones	10	Paris
L3	Blake	30	Paris
L4	Clark	20	London
L5	Adams	30	Athen

## Teile

Teil_Nr	Teil_Name	Teil_Farbe	Teil_Gewicht	Teil_Stadt
T1	Mutter	Rot	12	London
T2	Bolzen	Grün	17	Paris
T3	Schraube	Blau	17	Rom
T4	Schraube	Rot	14	London
T5	Nocken	Blau	12	Paris
T6	Zahnrad	Rot	19	London

## Projekte

Projekt Nr	Projekt Name	Projekt Stadt
P1	Sortierer	Paris
P2	Bildschirm	Rom
P3	OCR	Athen
P4	Konsole	Athen
P5	RAID	London
P6	EDS	Oslo
P7	Bandlaufwerk	London

## LTP – Lieferungen

LTP_Lieferant_Nr	LTP_Teil_Nr	LTP_Projekt_Nr	LTP_Menge
L1	T1	P1	200
L1	T1	P4	700
L2	T3	P1	400
L2	T3	P2	200
L2	T3	P3	200
L2	T3	P4	500
L2	T3	P5	600
L2	T3	P6	400
L2	T3	P7	800
L2	T5	P2	100
L3	T3	P1	200
L3	T4	P2	500
L4	T6	P3	300
L4	T6	P7	300
L5	T1	P4	100
L5	T2	P2	200
L5	T2	P4	100
L5	T3	P4	200
L5	T4	P4	800
L5	T5	P4	400
L5	T5	P5	500
L5	T5	P7	100
L5	T6	P2	200
L5	T6	P4	500