2. Übungsblatt

3.0 VU Datenmodellierung

8. Mai 2017

Allgemeines

In diesem Übungsteil sollen Sie Aufgabenstellungen aus den Bereichen SQL und Normalformentheorie bearbeiten.

Lösen Sie die Beispiele **eigenständig**, denn bei der Prüfung und in der Praxis sind Sie auch auf sich alleine gestellt. Wir weisen Sie darauf hin, dass abgeschriebene Lösungen mit 0 Punkten beurteilt werden.

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

Deadlines

bis 01.06.	**06:55 Uhr**	Upload über TUWEL
ab 08.06.	13:00 Uhr	Feedback in TUWEL verfügbar
bis 11.06.	23:59 Uhr	Reservierung eines Termins für das Tutorengespräch
bis 11.06.	23:59 Uhr	Reservierung eines Termins für das Abgabegespräch

(Terminreservierung für Tutoren- und Abgabegespräch über TUWEL)

Tutorengespräch

1. Sie müssen sich über TUWEL zu einem Tutorengespräch anmelden. Bitte machen Sie das rechtzeitig, je später Sie sich anmelden, umso eingeschränkter ist das Terminangebot.

2. Wenn Sie kein Blatt abgegeben haben, sind Sie nicht zum Tutorengespräch zugelassen.

3. Sie kommen mit Ihrem Studierendenausweis zu der von Ihnen reservierten Zeit vorbei, und absolvieren das Tutorengespräch. Sie erhalten dabei wertvolles Feedback zu Ihrem Übungsblatt. Das Gespräch ist verpflichtend und dauert ca. 30 Minuten.

4. Nützen Sie das Tutorengespräch um hilfreiches Feedback zu Ihrem Übungsblatt zu erhalten: Gehen Sie vor dem Gespräch Ihre korrigierte Lösungen durch, identifizieren Sie Unklarheiten und stellen Sie den TutorInnen entsprechende **Fragen**. Trauen Sie sich zuzugeben, wenn Sie etwas nicht verstanden haben. Ihre Fragen haben keinerlei negativen Einfluss auf Ihre Note!

Abgabegespräch

1. Sie müssen sich über TUWEL zu einem Abgabegespräch anmelden. Bitte machen Sie das rechtzeitig, je später Sie sich anmelden, umso eingeschränkter ist das Terminangebot.

2. Sie kommen mit Ihrem Studierendenausweis zu der von Ihnen reservierten Zeit vorbei, und absolvieren das Abgabegespräch. Stoffgebiet des Abgabegesprächs sind die mit dem Übungsblatt abgedeckten Themengebiete. Wir setzen voraus, dass Sie sich mit Ihrer korrigierten Abgabe auseinandergesetzt haben. Das Abgabegespräch hat Prüfungscharakter.

3. Sie absolvieren Ihr Abgabegespräch gemeinsam mit anderen KollegInnen. Das Gespräch dauert ca. 60 Minuten.

4. Sie können auf die Abgabe maximal 15 Punkte erreichen. Diese setzen sich wie folgt zusammen:

   5 Punkte auf das Übungsblatt

   10 Punkte auf das Abgabegespräch

5. Die Assistenten tragen die Punkte des Abgabegesprächs in TUWEL ein und Sie sehen dort, wieviele Punkte Sie bekommen haben.

6. Falls Sie nicht zu Ihrem Gesprächstermin erscheinen, bekommen Sie automatisch 0 Punkte auf das Gespräch.

SQL

Aufgabe 1 (eSQL) *[2.0 Punkte]*
Lösen Sie alle 10 (inkl. Unterpunkte) in unserer Online-Übungsumgebung zur Verfügung gestellten SQL-Aufgaben. Sie erreichen die Umgebung über TUWEL: Wählen Sie im Abschnitt "2.Übungsblatt" die Aktivität **eSQL Tool**. Sie benötigen kein weiteres Passwort, die Authentifizierung erfolgt über TUWEL.

Der verpflichtende SQL-Test am Ende des Semesters wird über die selbe Plattform abgewickelt. Es empfiehlt sich daher zusätzlich auch mit Beispielen aus den vorigen Semestern zu üben.

Normalformentheorie

Aufgabe 2 (Funktionale Abhängigkeiten) *[0.5 Punkte]*

(a) Gegeben ein Relationenschema $\mathcal{R} = ABCDE$ und die Mengen $\mathcal{F}_1$ und $\mathcal{F}_2$ von funktionalen Abhängigkeiten:

$$\mathcal{F}_1 = \{C \to DE,\ D \to B,\ AB \to BCD,\ AD \to CE,\ BE \to A,\ BD \to E\}$$
$$\mathcal{F}_2 = \{C \to D,\ DE \to BCE,\ AB \to DE,\ BE \to A\}$$

Sind $\mathcal{F}_1$ und $\mathcal{F}_2$ äquivalent? Begründen Sie Ihre Antwort!

Lösung:
Die Mengen von FDs $\mathcal{F}_1$ und $\mathcal{F}_2$ sind nicht äquivalent. Um dies zu zeigen reicht es aus, eine Menge an Attributen X zu finden, sodass $\mathit{AttrHülle}(\mathcal{F}_1, X) \neq \mathit{AttrHülle}(\mathcal{F}_2, X)$ gilt. Die Menge $\{D\}$ erreicht das Gewünschte:
Es gilt $\mathit{AttrHülle}(\mathcal{F}_1, D) = \{A, B, C, D, E\}$ und $\mathit{AttrHülle}(\mathcal{F}_2, \{D\}) = \{D\}$.

Obiger Ansatz zeigt, dass die FD $D \to ABCDE \in \mathcal{F}_1^+$, jedoch nicht in $\mathcal{F}_2^+$. Ein alternativer Ansatz wäre es, z.B. die FD $D \to B$ aus $\mathcal{F}_1$ herzunehmen und zu zeigen dass diese nicht in $\mathcal{F}_2^+$ liegt. Dies geschieht am leichtesten wiederum indem man $\mathit{AttrHülle}(\mathcal{F}_2, \{D\}) = \{D\}$ berechnet. Auf Grund des Zusammenhangs $D \to B \in \mathcal{F}_2^+$ (d.h., $\mathcal{F}_2 \models D \to B$) genau dann wenn $B \in \mathit{AttrHülle}(\mathcal{F}_2, \{D\})$ hat man damit gezeigt dass $\mathcal{F}_1$ und $\mathcal{F}_2$ nicht äquivalent sind.

(b) Gegeben das Schema $\mathcal{R} = ABCD$ der Relation R, auf der die Menge von funktionalen Abhängigkeiten $\mathcal{F} = \{A \to B,\ C \to D,\ BD \to A\}$ gültig ist. Weiters enthält die Relation R die Tupel $(0, 0, 0, 0)$ und $(1, 0, 1, 1)$. Welche weiteren Tupel aus der Menge $\{0, 1\}^4$ kann R enthalten, ohne eine FD in $\mathcal{F}$ zu verletzen?

Lösung:
Sei $(a, b, c, d) \in R \cap \{0, 1\}^4$.

- Falls $a = 0$, muss auch $b = 0$ sein, da sonst die FD $A \to B$ verletzt ist. Somit müssen wir die Fälle $c = 0$ und $c = 1$ unterscheiden. Falls $c = 0$, muss auch $d = 0$ sein, da sonst die FD $C \to D$ verletzt wird. Damit erhält man das bereits bekannte Tupel $(a, b, c, d) = (0, 0, 0, 0)$. Falls $c = 1$, muss auch $d = 1$ sein, da wir sonst die FD $C \to D$ verletzen. Die Wahl von $a = 0, b = 0, c = 1, d = 1$ verletzt aber die FD $BD \to A$.
- Falls $a = 1$, muss auch $b = 0$ sein, da sonst $A \to B$ verletzt wird. Deswegen muss $d = 1$ gelten, da sonst $BD \to A$ verletzt wird. Das hat zur Folge, dass $c = 1$ gelten muss, da sonst $C \to D$ verletzt wird, womit wir das bereits bekannte Tupel $(a, b, c, d) = (1, 0, 1, 1)$ erhalten.

Insgesamt kann R neben $(0, 0, 0, 0)$ und $(1, 0, 1, 1)$ kein anderes $\{0, 1\}$-Tupel enthalten, ohne die FDs zu verletzen.

Anmerkung: Ein anderer, jedoch umständlicher Lösungsweg ist, für alle 14 Lösungskandidaten zu untersuchen, ob sie die FDs verletzen.

Aufgabe 3 (Kanonische Überdeckung) *[0.5 Punkte]*

Bestimmen Sie eine kanonische Überdeckung folgender Mengen funktionaler Abhängigkeiten über dem Relationenschema $\mathcal{R}$ und dokumentieren Sie den Lösungsweg.

(a) $\mathcal{R} = ABCDEF$,
$\mathcal{F}_1 = \{A \to C,\ E \to ABF,\ AB \to F,\ AF \to C,\ BC \to A,\ CDE \to A\}$

(b) $\mathcal{R} = ABCDEF$,
$\mathcal{F}_2 = \{D \to AC,\ E \to BD,\ AD \to ACE,\ AE \to BEF,\ CF \to ABD\}$

Lösung:

(a) Die Kanonische Überdeckung lässt sich mithilfe der folgenden vier Schritten berechnen:

- Dekomposition:

  Mit der Regel der Dekomposition erhalten wir die Menge
  $\{A \to C,\ E \to A,\ E \to B,\ E \to F,\ AB \to F,\ AF \to C,\ BC \to A,\ CDE \to A\}$.

- Linksreduktion:

  Wir müssen uns jetzt sukzessive die Frage stellen, ob für eine funktionale Abhängigkeit ein Attribut auf der linken Seite überflüssig ist. Bei $AF \to C$ kann F gestrichten werden, da bereits $A \to C$ gilt, und bei $CDE \to A$ können sowohl C als auch D gestrichen werden, da $E \to A$ gilt. Somit ist die Menge der funktionalen Abhängigkeiten nach der Linksreduktion (und Streichung von doppelten FDs) gegeben als
  $\{A \to C,\ E \to A,\ E \to B,\ E \to F,\ AB \to F,\ BC \to A\}$.

- Rechtsreduktion:

  Bei der Rechtsreduktion muss man für jede funktionale Abhängigkeit überprüfen, ob sie noch gültig ist, selbst wenn man sie streicht. Für $E \to F$ ist das hier der Fall: Aus E folgt A und auch B, Darum gilt $E \to AB$. Somit folgt aus $AB \to F$ und Transitivität dann $E \to F$, ohne die FD $E \to F$ zu verwenden. Die FD $E \to F$ kann also gestrichen werden, und es ist hier auch die einzige, die gestrichen werden muss. Somit ist die Menge der funktionalen Abhängigkeiten nach der Linksreduktion und Vereinigung gegeben als
  $\{A \to C,\ E \to AB,\ AB \to F,\ BC \to A\}$, was der kanonischen Überdeckung entspricht.

(b)
- Dekomposition:
  $\{D \to A,\ D \to C,\ E \to B,\ E \to D,\ AD \to A,\ AD \to C,\ AD \to E,\ AE \to B,\ AE \to E,\ AE \to F,\ CF \to A,\ CF \to B,\ CF \to D\}$

- Linksreduktion:

  Die FD $AD \to A$ kann man entweder zuerst linksreduzieren (**Achtung**: Hier wird D gestrichen, und nicht A, also $AD \to A$ kann zu $A \to A$ reduziert werden) und danach rechtsreduzieren, oder einfach sofort streichen. Im Allgemeinen gilt: Wenn eine FD trivial ist, kann sie gestrichen werden; sie hat keinen Einfluss auf das Ergebnis der kanonischen Überdeckung.
  Bei $AD \to C$ kann A gestrichen werden, da bereits $D \to C$ gilt. Auch bei $AD \to E$ kann man A streichen: Aus $D \to A$ gilt mit Verstärkung $D \to AD$, und da $AD \to E$ gilt folgt aus Transitivität auch $D \to E$. Nach weiteren Reduktionsschritten und Streichung von doppelten FDs erhält man nach der Linksreduktion somit

  $\{D \to A,\ D \to C,\ E \to B,\ E \to D,\ D \to E,\ E \to F,\ CF \to A,\ CF \to B,\ CF \to D\}$

- Rechtsreduktion

  Die FD $D \to A$ kann gestrichen werden: Aus D folgt E und C und aus E folgt F, somit gilt $D \to CF$. Damit gilt mit $CF \to A$ auch $D \to A$. Da also implizit $D \to A$ aus den anderen FDs folgt, kann sie gestrichen werden. Durch weitere Streichungen und Zusammenfassen erhalten wir folglich:

  $(\mathcal{F}_2)_C = \{D \to CE,\ E \to DF,\ CF \to ABD\}$

Aufgabe 4 (Schlüsselbestimmung) *[0.5 Punkte]*

Bestimmen Sie für die folgenden Relationenschemata samt funktionalen Abhängigkeiten alle Schlüssel und alle Superschlüssel.

(a)

$$\mathcal{R} = ABCD$$
$$\mathcal{F} = \{BC \to A,\ AD \to B,\ CD \to B,\ AC \to D\}$$

Lösung:
Die Schlüssel sind BC, CD und AC. Die Menge der Superschlüssel:

$$\{AC, BC, CD, ABC, ACD, BCD, ABCD\}.$$

(b)

$$\mathcal{R} = ABCDEFG$$
$$\mathcal{F} = \{A \to C,\ BD \to E,\ DF \to C,\ ABC \to DE,\ CDE \to AF\}$$

Lösung:
Die Schlüssel sind ABG, $BCDG$ und $BDFG$. Die Menge der Superschlüssel:

$$\{ABG, ABCG, ABDG, BCDG, ABEG, ABFG, BDFG, ABCDG, ABCEG, ABDEG, BCDEG, ABCFG, ABDFG, BCDFG, ABEFG, BDEFG, ABCDEG, ABCDFG, ABCEFG, ABDEFG, BCDEFG, ABCDEFG\}$$

Aufgabe 5 (Normalformen) *[0.5 Punkte]*

Gegeben sei ein Relationenschema $\mathcal{R}$ samt dazugehörigen funktionalen Abhängigkeiten. Geben Sie an, ob $\mathcal{R}$

- in dritter Normalform ist,
- in Boyce-Codd-Normalform ist,

und begründen Sie Ihre Antworten.

(a)

$$\mathcal{R} = ABCDEF$$
$$\mathcal{F} = \{E \to BE,\ AB \to D,\ AD \to AE,\ BF \to A,\ CD \to AF,\ DF \to ABC\}$$

Lösung:
Die Schlüssel sind CD, BF, DF, EF, ABC und ACE. Die drei Anforderungen für funktionale Abhängigkeiten der Form $\alpha \to B$, die man benötigt um ein Relationenschema $\mathcal{R}$ auf dritte Normalform bzw. Boyce-Codd Normalform zu untersuchen, sind wie folgt:

1. $B \in \alpha$, d.h. die FD ist trivial
2. α ist Superschlüssel von $\mathcal{R}$
3. das Attribut B ist in einem Schlüssel von $\mathcal{R}$ enthalten

Dabei ist ein Relationenschema $\mathcal{R}$ in Boyce-Codd Normalform, wenn für jede funktionale Abhängigkeit Bedingungen 1 oder 2 erfüllt ist, und $\mathcal{R}$ ist in dritter Normalform wenn für jede funktionale Abhängigkeit Bedingungen 1,2 oder 3 erfüllt ist. Um dies zu untersuchen wenden wir die Dekomposition an, und kennzeichnen für jede FD, welche Bedingungen sie erfüllt:

$$\mathcal{F} = \{\underbrace{E \to B}_{3}, \underbrace{E \to E}_{1,3}, \underbrace{AB \to D}_{3}, \underbrace{AD \to A}_{1,3}, \underbrace{AD \to E}_{3}, \underbrace{BF \to A}_{2,3},$$
$$\underbrace{CD \to A}_{2,3}, \underbrace{CD \to F}_{2,3}, \underbrace{DF \to A}_{2,3}, \underbrace{DF \to B}_{2,3}, \underbrace{DF \to C}_{2,3}\}.$$

Für jede FD trifft entweder Bedingung 1,2 oder 3 zu. Zudem gibt es mindestens eine FD (z.B. $E \to B$), für die Bedingung 1 und 2 nicht zutreffen. Deswegen ist $\mathcal{R}$ in dritter Normalform, aber nicht in BCNF.

(b)

$$\mathcal{R} = ABCDEFG$$
$$\mathcal{F} = \{FB \to A,\ F \to BCD,\ B \to EFG\}$$

Lösung:
Die Schlüssel von $\mathcal{R}$ sind B und F. Auf der linken Seite jeder FD steht daher ein Superschlüssel, damit ist $\mathcal{R}$ in BCNF.

Aufgabe 6 (Synthesealgorithmus) *[0.5 Punkte]*

Gegeben sei folgendes Relationenschema samt funktionalen Abhängigkeiten:

$$\mathcal{R} = ABCDE$$
$$\mathcal{F} = \{AD \to BC,\ CE \to A,\ ACD \to B\}$$

Gesucht ist eine verlustlose und abhängigkeitserhaltende Zerlegung in dritter Normalform. Wenden Sie hierzu den Synthesealgorithmus an und dokumentieren Sie das Ergebnis der einzelnen Schritte. Bestimmen Sie alle Schlüssel von $\mathcal{R}$ und allen Relationen der Zerlegung.

Lösung:

1. Bestimmung der kanonischen Überdeckung:

$$\mathcal{F}_c = \{AD \to BC,\ CE \to A\}$$

2. Erstelle Relationenschemata für jedes Element von $\mathcal{F}_c$:

Relationenschema	Geltende FDs
$\mathcal{R}_1 = ABCD$	$\mathcal{F}_1 = \{AD \to BC\}$
$\mathcal{R}_2 = ACE$	$\mathcal{F}_2 = \{CE \to A\}$

3. Bestimmung aller Kandidatenschlüssel von $\mathcal{R}$ bzgl. $\mathcal{F}_c$: ADE, CDE.

4. Keiner der Schlüssel ist in einem Teilschema anthalten. Es muss ein neues Relationenschema $\mathcal{R}_3$ erstellt werden:

Relationenschema	Geltende FDs
$\mathcal{R}_3 = ADE$	$\mathcal{F}_3 = \emptyset$

5. Es muss kein Schema eliminiert werden.

Ergebnis (Schlüssel sind unterstrichen):

Relationenschema	Geltende FDs
$\mathcal{R}_1 = \underline{AD}BC$	$\mathcal{F}_1 = \{AD \to BC\}$
$\mathcal{R}_2 = A\underline{CE}$	$\mathcal{F}_2 = \{CE \to A\}$
$\mathcal{R}_3 = \underline{ADE}$	$\mathcal{F}_3 = \emptyset$

Aufgabe 7 (Dekompositionsalgorithmus) *[0.5 Punkte]*

Gegeben sei folgendes Relationenschema samt funktionalen Abhängigkeiten:

$$R = UVWXY$$
$$\mathcal{F} = \{U \to VW,\ W \to X,\ VY \to X\}$$

Gesucht ist eine verlustlose Zerlegung in Boyce-Codd-Normalform. Wenden Sie hierzu den Dekompositionsalgorithmus an und dokumentieren Sie das Ergebnis der einzelnen Schritte. Bestimmen Sie alle Schlüssel von $\mathcal{R}$ und allen Relationen der Zerlegung. Ist die Zerlegung abhängigkeitserhaltend? Wenn die Zerlegung nicht abhängigkeitserhaltend ist, geben Sie an, welche Abhängigkeiten verloren gegangen sind. *Hinweis*: Bestimmen Sie bei jeder Zerlegung die jeweilige Hülle an FDs!

Lösung:
Der einzige Schlüssel von $\mathcal{R}$ is UY. Alle drei FDs verletzen die BCNF, da sie nicht trivial sind, und keine der Attributmengen U, W oder VY ein Superschlüssel ist. Daher gibt es drei Möglichkeiten für eine Zerlegung:

Hinweis: Wir geben in jeder Hülle $\mathcal{F}_i^+[\mathcal{R}_j]$ nur die Nicht-trivialen Abhängigkeiten an.

- Die FD $U \to VW$ wird gewählt. Damit erhalten wird:

$$\begin{array}{lll} \mathcal{R}_1 = UVW & \mathcal{F}_1 = \mathcal{F}^+[\mathcal{R}_1] = \{U \to VW\} & \text{Schlüssel: } U \\ \mathcal{R}_2 = UXY & \mathcal{F}_2 = \mathcal{F}^+[\mathcal{R}_2] = \{U \to X\} & \text{Schlüssel: } UY \end{array}$$

  R_2 ist nicht in BCNF, muss also weiter zerlegt werden zu:

$$\begin{array}{lll} \mathcal{R}_1 = UVW & \mathcal{F}_1 = \mathcal{F}^+[\mathcal{R}_1] = \{U \to VW\} & \text{Schlüssel: } U \\ \mathcal{R}_{2,1} = UX & \mathcal{F}_{2,1} = \mathcal{F}_2^+[\mathcal{R}_{2,1}] = \{U \to X\} & \text{Schlüssel: } U \\ \mathcal{R}_{2,2} = UY & \mathcal{F}_{2,2} = \mathcal{F}_2^+[\mathcal{R}_{2,2}] = \emptyset & \text{Schlüssel: } UY \end{array}$$

  Diese Zerlegung ist nicht abhängikeitserhaltend, da $VY \to X$ verloren geht.

- Die FD $W \to X$ wird gewählt. Damit erhalten wird:

$$\begin{array}{lll} \mathcal{R}_1 = WX & \mathcal{F}_1 = \mathcal{F}^+[\mathcal{R}_1] = \{W \to X\} & \text{Schlüssel: } W \\ \mathcal{R}_2 = UVWY & \mathcal{F}_2 = \mathcal{F}^+[\mathcal{R}_2] = \{U \to VW\} & \text{Schlüssel: } UY \end{array}$$

R_2 ist nicht in BCNF, muss also weiter zerlegt werden zu:

$$\begin{array}{rcl} \mathcal{R}_1 = WX & \mathcal{F}_1 = \mathcal{F}^+[\mathcal{R}_1] = \{W \to X\} & \text{Schlüssel: } W \\ \mathcal{R}_{2,1} = UVW & \mathcal{F}_{2,1} = \mathcal{F}_2^+[\mathcal{R}_{2,1}] = \{U \to VW\} & \text{Schlüssel: } U \\ \mathcal{R}_{2,2} = UY & \mathcal{F}_{2,2} = \mathcal{F}_2^+[\mathcal{R}_{2,2}] = \emptyset & \text{Schlüssel: } UY \end{array}$$

Diese Zerlegung ist nicht abhängikeitserhaltend, da $VY \to X$ verloren geht.

- Die FD $VY \to X$ wird gewählt. Daraus ergibt sich

$$\begin{array}{rcl} \mathcal{R}_1 = VXY & \mathcal{F}_1 = \mathcal{F}^+[\mathcal{R}_1] = \{VY \to X\} & \text{Schlüssel: } VY \\ \mathcal{R}_2 = UVWY & \mathcal{F}_2 = \mathcal{F}^+[\mathcal{R}_2] = \{U \to VW\} & \text{Schlüssel: } UY \end{array}$$

R_2 ist nicht in BCNF, muss also weiter zerlegt werden zu:

$$\begin{array}{rcl} \mathcal{R}_1 = VXY & \mathcal{F}_1 = \mathcal{F}^+[\mathcal{R}_1] = \{VY \to X\} & \text{Schlüssel: } VY \\ \mathcal{R}_{2,1} = UVW & \mathcal{F}_{2,1} = \mathcal{F}_2^+[\mathcal{R}_{2,1}] = \{U \to VW\} & \text{Schlüssel: } U \\ \mathcal{R}_{2,2} = UY & \mathcal{F}_{2,2} = \mathcal{F}_2^+[\mathcal{R}_{2,2}] = \emptyset & \text{Schlüssel: } UY \end{array}$$

Diese Zerlegung ist nicht abhängikeitserhaltend, da $W \to X$ und auch $U \to X$ verloren gehen ($U \to X \in \mathcal{F}^+$!).