
Introduction to Data and Knowledge Engineering Lösungsvorschlag zu Tutorium 3



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Welche der folgenden Aussagen ist richtig? Begründen Sie jede Ihrer Antworten oder geben Sie ein Gegenbeispiel an!

In welcher Beziehung stehen Normalformen zueinander?

- Ist eine Relation in 3 NF, so ist sie auch in 2 NF.
- Ist eine Relation in 2 NF, so ist sie auch in 3 NF.
- 2 NF und 3 NF bedingen sich überhaupt nicht gegenseitig: Ist eine Relation in 2 NF, so kann sie auch in 3 NF sein, ist eine Relation in 3 NF, so kann sie auch in 2 NF sein, muss es aber nicht.

Die verschiedenen Normalformen bauen aufeinander auf; eine höhere Normalform bedeutet stets, dass die niedrigeren Normalformen auch eingehalten sein müssen. Umgekehrt gilt dies nicht unbedingt, da höhere Normalformen höhere Anforderungen haben (siehe zum Beispiel letzte Aufgabe).

Multiple Choice II

Wann ist eine Relation in 1 NF?

- Eine Relation ist immer in 1 NF.
- Eine Relation ist nicht immer in 1 NF.
- Eine Relation ist in 1 NF, wenn jedes Attribut der Relation einen atomaren Inhalt hat.

Was bedeutet die Aussage „alle Nichtschlüsselattribute hängen vollständig vom Primärschlüssel ab“?

- Jedes Attribut, welches nicht der Schlüssel einer Relation ist, hängt von mindestens einem Teil des Primärschlüssels dieser Relation ab.
- Kein Attribut, welches nicht der Schlüssel einer Relation ist, hängt ausschließlich von einem Teil des Primärschlüssels dieser Relation ab.
- Jedes Attribut, welches nicht der Schlüssel einer Relation ist, ist nicht transitiv vom Primärschlüssel dieser Relation abhängig.
- Jedes Attribut, welches nicht der Schlüssel einer Relation ist, ist von allen Schlüsselns dieser Relation abhängig.

Vollständige Abhängigkeit bedeutet, dass ein Attribut vom **gesamten** Primärschlüssel abhängen muss. Diese Abhängigkeit darf aber auch transitiv vorliegen.

Wann ist eine Relation in 2 NF?

- Wenn der Primärschlüssel einer Relation aus genau einem Attribut besteht, von dem alle anderen Attribute abhängig sind, ist die Relation in 2 NF.
- Wenn eine Relation höchstens so viele Attribute wie Schlüssel besitzt, ist sie in 2 NF.
- Wenn alle Nichtschlüsselattribute einer Relation vollständig vom Primärschlüssel abhängen, ist diese Relation in 2 NF.

Aussage 1 ist trivialerweise erfüllt: Besteht der Schlüssel nur aus einem Attribut, kann auch keine partielle Abhängigkeit vonseiten der Nichtschlüsselattribute vorliegen. Besteht der Schlüssel allerdings aus mehreren Attributen, muss auf die vollständige Abhängigkeit geachtet werden (siehe Definition von 2 NF).

Wann ist eine Relation in 3 NF?

- Wenn der Primärschlüssel einer Relation aus genau einem Attribut besteht, von dem alle anderen Attribute der Relation abhängig sind, ist die Relation in 3 NF.
- Wenn eine Relation höchstens ein Nichtschlüsselattribut besitzt, welches vollständig vom Primärschlüssel abhängig ist, ist sie in 3 NF.
- Wenn eine Relation in 2 NF ist und kein Nichtschlüsselattribut ausschließlich von anderen Nichtschlüsselattributen abhängig ist, ist die Relation in 3 NF.

Aussage 1 ist falsch, da transitive Abhängigkeit in der Aussage nicht ausgeschlossen wird. Aussage 2 ist wahr, denn falls es nur ein Nichtschlüsselattribut gibt, welches vollständig vom Primärschlüssel abhängig ist, ist 2 NF erfüllt und es kann keine weitere transitive Abhängigkeit vorliegen. Aussage 3 schließt transitive Abhängigkeit aus und fordert 2 NF - dies entspricht der Definition von 3 NF.



Was ist eine Determinant (im Datenbankenbereich)?

- Ein Determinant ist ein besonders wichtiges Attribut in einer Relation.
- Ein Determinant ist eine Attributmenge, von der irgendein anderes Attribut vollständig abhängig ist.
- Ein Determinant ist ein anderer Begriff für Schlüsselkandidat einer Relation.

Aussage 2 entspricht der Definition von Determinant. Ein Determinant muss zunächst nicht automatisch ein Schlüsselkandidat sein!

Wann ist eine Relation in BCNF?

- Wenn eine Relation in 3 NF ist, ist sie auch in BCNF.
- Wenn eine Relation in 3 NF ist und jeder Determinant ein Schlüsselkandidat der Relation ist, ist sie in 3 NF.
- Wenn eine Relation in 3 NF ist und nur einen Determinant besitzt, ist sie in 3 NF.

BCNF verschärft 3 NF - nicht jede Relation, die in 3 NF ist, ist daher zwingenderweise auch in BCNF (oft ist dies jedoch der Fall). Aussage 2 entspricht der Definition von BCNF. Aussage 3 ist wahr, denn besitzt eine Relation nur einen Determinant, muss dieser der Primärschlüssel der Relation sein. Da es keinen anderen Determinant gibt und zudem gefordert wird, dass die Relation in 3 NF ist, muss sie in diesem Fall auch in BCNF sein.

1. Normalform

Aufgabenstellung

Ein Hotel in einem Schweizer Skigebiet sucht einen Mitarbeiter / eine Mitarbeiterin zur Besetzung der Rezeption. Die Daten der bisherigen Bewerber werden im Augenblick folgendermaßen gespeichert (Schlüsselattribute sind in grau gekennzeichnet, keine Fremdschlüssel):

ID	Vorname	Nachname	Geburtsdatum	Sprachen	Adresse
001	Anna	Müller	4.5.1980	Deutsch, Englisch, Französisch	Beethovenweg 83, 12345 Großstadt
002	Werner	Hagen	6.7.1975	Deutsch, Englisch	Hauptstraße 5, 34342 Kleindorf
003	Claire	Calmet	12.12.1989	Französisch	Boulevard Wilson 33, 06600 Antibes
004	Lisa	Schmidt	1.4.1977	Deutsch, Englisch, Spanisch	Berliner Straße 19, 45671 Anderstadt
005	Harry	Genius	4.3.1990	Deutsch, Englisch, Französisch, Spanisch, Schwedisch, Finnisch, Russisch	Bessunger Straße 44, 64285 Darmstadt

Ist das Modell in 1 NF? Falls nein, wie könnte es verändert werden, damit 1 NF erfüllt wird? Tragen Sie ggf. zur Veranschaulichung auch die gegebenen Beispieldaten in Ihr verändertes Relationenschema ein.

1. Normalform erfüllt?

Dieses Relationenschema ist nicht in 1 NF, da in der Spalte „Sprachen“ und in der Spalte „Adresse“ keine atomaren Werte gespeichert sind: In der Spalte „Sprachen“ sind teilweise mehrere verschiedene Sprachen gespeichert. In der Spalte „Adresse“ sind 4 unterschiedliche Adressinformationen gespeichert (Straße, Hausnummer, Postleitzahl und Stadt).

Probleme, die dadurch auftreten:

- ▶ Verschiedene Informationen werden in einem Attribut gespeichert statt in verschiedenen - eine einzelne solche Information kann also nur schwer abgefragt werden.
- ▶ In einem Attribut wird eine Menge von Attributwerten gespeichert - auch hierdurch ist die Abfrage einzelner Werte nicht möglich (beispielsweise kann man per Datenbankabfrage nicht einfach herausfinden, wer alles Englisch spricht).

1. Normalform

Lösungsansatz

1. Möglichkeit: Jede der betroffenen Spalten aufteilen. Bei der Spalte „Adresse“ ist das noch recht einfach:

Straße	No	PLZ	Stadt
Beethovenweg	83	12345	Großstadt

usw.

Die Spalte „Sprache“ könnte man folgendermaßen aufteilen:

Sprache 1	Sprache 2	Sprache 3	...
Deutsch	Englisch	Französisch	...

Problem: Wie viele Spalten sollen für Sprachen vorgesehen werden?

Die Maximalzahl der Sprachen, die ein Bewerber sprechen kann, ist hier nicht bekannt.

1. Normalform

Besserer Lösungsansatz

2. Möglichkeit: Sprache wird zusätzlich zur Bewerber-ID zum Schlüsselattribut und für jeden Bewerber und jede von ihm gesprochene Sprache wird eine Zeile angelegt; die Spalte „Adresse“ wird aufgeteilt wie vorher vorgeschlagen:

ID	Vorname	Nachname	Geb.datum	Sprachen	Straße	No	PLZ	Stadt
001	Anna	Müller	4.5.1980	Deutsch	Beethovenweg	83	12345	Großstadt
001	Anna	Müller	4.5.1980	Englisch	Beethovenweg	83	12345	Großstadt
001	Anna	Müller	4.5.1980	Französisch	Beethovenweg	83	12345	Großstadt
002	Werner	Hagen	6.7.1975	Deutsch	Hauptstraße	5	34342	Kleindorf
002	Werner	Hagen	6.7.1975	Englisch	Hauptstraße	5	34342	Kleindorf

usw.

Nun können für jeden Bewerber so viele Sprachen eingefügt werden wie nötig.
Nachteil: Viele Redundanzen; einige Informationen werden mehrfach abgespeichert (siehe nächste Aufgabe).

2. Normalform Aufgabenstellung I

Ein Fachgebiet bietet verschiedene Lehrveranstaltungen an. Jede dieser Lehrveranstaltungen hat einen Namen (VName) und einen eindeutigen Schlüssel (VID), welcher der Veranstaltung vom Prüfungssekretariat des Fachbereichs zugewiesen wurde. Das Fachgebiet erlaubt in jeder seiner Veranstaltungen generell, dass Gruppen von Studenten gemeinsam die Übungen bearbeiten. Die maximale Größe dieser Gruppen wird zunächst nicht festgelegt, da sie in jeder Veranstaltung anders sein kann. Jede dieser Gruppen bekommt hierfür eine innerhalb der Veranstaltung eindeutige Gruppennummer (GID). Studenten besitzen außerdem jeweils eindeutige Matrikelnummern (Matr), denen wiederum Vornamen und Nachnamen zugeordnet sind. Ein Student kann mehrere der angebotenen Lehrveranstaltungen besuchen, aber innerhalb einer Lehrveranstaltung nur einer Gruppe zugeordnet sein. In einer Datenbank sollen diese Zuordnungen gespeichert werden. In einem ersten Versuch entstand folgendes Relationen-Schema (Schlüsselattribute sind in grau gekennzeichnet, keine Fremdschlüssel), für das zur Veranschaulichung einigen Beispiel-Entitäten angegeben sind:

2. Normalform Aufgabenstellung II

VID	VName	GID	Matr	Vorname	Nachname
345	Angewandte Physik 1	1	1234567	Arno	Schmidt
345	Angewandte Physik 1	1	7654321	Anna	Müller
345	Angewandte Physik 1	2	3344559	Helmut	Groß
345	Angewandte Physik 1	2	3366771	Kevin	Amundsen
675	Theoretische Physik 1	1	5829455	Jacqueline	Herr
675	Theoretische Physik 1	1	3452649	Angela	Bennett
675	Theoretische Physik 1	2	1234567	Arno	Schmidt
675	Theoretische Physik 1	2	3344559	Helmut	Groß

Welche Probleme können hier auftreten? Was kann beispielsweise passieren, wenn Jacqueline sich jetzt auch in eine Übungsgruppe der Veranstaltung Angewandte Physik 1 eintragen lassen möchte?

Ist das Modell (d.h. das Relationen-Schema zusammen mit den Funktionalen Abhängigkeiten) in 2 NF? Falls nein, wie müsste es verändert werden, damit 2 NF erfüllt wird? Tragen Sie ggf. zur Veranschaulichung auch die gegebenen Beispieldaten in Ihr verändertes Relationenschema ein.

2. Normalform erfüllt?

2 NF ist in diesem Beispiel nicht erfüllt, da nicht alle Attribute vollständig funktional vom Primärschlüssel abhängig sind:

- ▶ VName ist nur von VID abhängig, aber nicht von Matr
- ▶ Vorname und Nachname der Studenten ist nur von Matr abhängig, aber nicht von VID
- ▶ Nur GID ist sowohl von VID als auch von Matr abhängig

Welche Probleme können hier auftreten?

Update-Anomalie

Ist eine Relation nicht in 2 NF, entstehen Datenredundanzen: Manche Informationen werden unnötigerweise mehrfach abgespeichert (im vorliegenden Beispiel vor allem der Name der Veranstaltung, Vor- und Nachname der Studenten). Dadurch können Inkonsistenzen entstehen, wenn ein Tupel der Relation aktualisiert werden soll oder ein neues hinzugefügt werden soll.

Wenn Jacqueline Herr sich beispielsweise in eine Übungsgruppe in Angewandte Physik 1 eintragen lassen möchte, könnten unter anderem Fehler beim Ausfüllen der Spalten VName, Vorname oder Nachname entstehen, wodurch die Informationen in der Datenbank nicht mehr konsistent sind:

VID	VName	GID	Matr	Vorname	Nachname
345	Angewandte Physik 1	1	1234567	Arno	Schmidt
345	Angewandte Physik 1	1	7654321	Anna	Müller
345	Angewandte Physik 1	2	3344559	Helmut	Groß
345	Angewandte Physik 1	2	3366771	Kevin	Amundsen
345	Angewandte Physik 2	3	5829455	Jenny	Herr
675	Theoretische Physik 1	1	5829455	Jacqueline	Herr
675	Theoretische Physik 1	1	3452649	Angela	Bennett
675	Theoretische Physik 1	2	1234567	Arno	Schmidt
675	Theoretische Physik 1	2	3344559	Helmut	Groß

2. Normalform

Lösungsvorschlag



VID	VName
345	Angewandte Physik 1
675	Theoretische Physik 1

Matr	Vorname	Nachname
1234567	Arno	Schmidt
7654321	Anna	Müller
3344559	Helmut	Groß
3366771	Kevin	Amundsen
5829455	Jacqueline	Herr
3452649	Angela	Bennett

VID	Matr	GID
345	1234567	1
345	7654321	1
345	3344559	2
345	3366771	2
675	5829455	1
675	3452649	1
675	1234567	2
675	3344559	2

3. Normalform Aufgabenstellung I

Eine Supermarktkette besitzt Filialen ganz Deutschland, von denen jede eine eindeutige Filialnummer (FID) hat. Für jede Filiale wird die Adresse gespeichert (Straße, Hausnummer, PLZ und Stadt) sowie der Nettoeinkommensdurchschnitt der Einwohner, welche in dem PLZ-Bereich wohnen, in dem sich die Filiale befindet (E-AVG). Mit dieser Information soll das Warenangebot in der entsprechenden Filiale reguliert werden. Bisher werden diese Informationen in folgendem Relationen-Schema (Schlüsselattribute sind in grau gekennzeichnet, keine Fremdschlüssel), das mit hier einigen Beispiel-Instanzen gefüllt ist:

3. Normalform Aufgabenstellung II

FID	Straße	Haus	PLZ	Stadt	E-AVG
421	Hauptstraße	12	45632	Altstadt	1 500
326	Nebengasse	42	56723	Armestadt	450
798	Große Straße	63	40195	Mittelstadt	2 000
667	Kleine Straße	55	40195	Mittelstadt	2 000
592	Schöne Straße	4	59603	Reichestadt	10 000
533	Lindenstraße	1	55555	Reichestadt	9 000
444	Lange Straße	1044	59603	Reichestadt	10 000
495	Großer Weg	230	34896	Halbestadt	3 000
245	Mittelweg	45	46245	Anderestadt	4 000
405	Villenstraße	7	40696	Millonärsstadt	2 000 000

Was ist an dem Relationenschema nicht optimal?

Ist das Modell in 3 NF? Falls nein, wie müsste es verändert werden, damit 3 NF erfüllt wird? Tragen Sie ggf. zur Veranschaulichung auch die gegebenen Beispieldaten in Ihr verändertes Relationenschema ein.

Überprüfen Sie anschließend, ob Ihre Lösung auch BCNF erfüllt.

3. Normalform erfüllt?

3 NF ist hier nicht erfüllt, da E-AVG und Stadt hier zunächst nur von PLZ abhängen; da PLZ von FID abhängt, hängen E-AVG und Stadt transitiv von FID ab. Auch hier sind wieder Datenredundanzen enthalten: Wird eine neue Filiale in einem PLZ-Gebiet angelegt, müssen für dieses PLZ-Gebiet erneut der Stadtname und das durchschnittliche Einkommen der Einwohner (E-AVG) eingetragen werden. Auch hierbei können wieder Inkonsistenzen entstehen:

FID	Straße	Haus	PLZ	Stadt	E-AVG
798	Große Straße	63	40195	Mittelstadt	2 000
667	Kleine Straße	55	40195	Mittelstadt	2 000
400	Mittelstraße	67	40195	Mittelstadt	1 500

3. Normalform

Lösungsvorschlag



FID	Straße	Haus	PLZ
421	Hauptstraße	12	45632
326	Nebengasse	42	56723
798	Große Straße	63	40195
667	Kleine Straße	55	40195
592	Schöne Straße	4	59603
533	Lindenstraße	1	55555
444	Lange Straße	1044	59603
495	Großer Weg	230	34896
245	Mittelweg	45	46245
405	Villenstraße	7	40696

PLZ	Stadt	E-AVG
45632	Altstadt	1 500
56723	Armestadt	450
40195	Mittelstadt	2 000
59603	Reichstadt	10 000
55555	Reichstadt	9 000
59603	Reichstadt	10 000
34896	Halbestadt	3 000
46245	Anderestadt	4 000
40696	Millonärsstadt	2 000 000

BCNF erfüllt?

BCNF ist erfüllt, da es nun innerhalb der Relationen keine weiteren Determinanten mehr gibt: Keine andere Attributmenge außer dem Primärschlüssel der jeweiligen Relation bestimmt irgendeine andere Attributmenge.

BCNF soll verhindern, dass Teile eines Schlüssels, der aus mehreren Attributen besteht, gegenseitig voneinander abhängig sind.