

Informatik

Datenbanken

5. Klasse TFO Brixen

Michael Mutschlechner

-
- ▶ PROBLEME NICHT ADÄQUATER MODELLIERUNG - ANOMALIEN
 - ▶ NORMALFORMEN



Normalisierung

- ▶ Nach der Anforderungsdefinition, dem ER-Modell und dem relationalen Datenmodell haben wir einen Datenbankentwurf. Nun folgt die Normalisierung des Datenbankentwurfs.
- ▶ Was versteht man unter Normalisierung?
 - ▶ Dazu das Beispiel des „Schwarzen Brettes“.
 - ▶ Wir nehmen folgende Datenstruktur an: Alle Daten seien in einer Tabelle
 - ▶ Legende: Anz.Nr.: Anzeigennummer, Ins.Nr.: Inserentennummer, Rub.Nr.: Rubriknummer



Beispiel

| <u>Anz.Nr.</u> | Anzeigentext, -datum | Ins.Nr. | Nickname, E-Mail | <u>Rub.Nr.</u> | Rubrikbezeichnung |
|----------------|------------------------------------|---------|---------------------------|----------------|-------------------|
| 501 | Mountainbike, 15.03.2016 | 214 | Sissi sissi@jker.de | 2 | Zweiräder |
| 502 | 5 Kinderbücher, 15.03.2016 | 211 | Franki franki@qwert.it | 5 | Bücher |
| 505 | Harry Potter 5, 14.03.2016 | 211 | Franki franki@qwert.it | 5 | Bücher |
| 507 | Buch TCP/IP Grundl., 02.03.2016 | 210 | Jens jens@wert.au | 3 | Computer |
| 507 | Buch TCP/IP Grundl., 02.03.2016 | 210 | Jens jens@wert.au | 5 | Bücher |
| 507 | Buch TCP/IP Grundl., 02.03.2016 | 210 | Jens jens@wert.au | 6 | Sonstiges |
| 508 | Gebr. Trekkingrad, 02.03.2016 | 214 | Sissi sissi@jker.de | 2 | Zweiräder |



Redundanz und Anomalien

- ▶ Redundante Informationen sind Informationen, die mehrfach vorkommen
 - ▶ Beispiel: Anzeigentext zu Anzeigennummer 507 kommt dreimal vor...
 - ▶ Speicherplatz wird verschwendet
- ▶ Redundanzen führen außerdem zu Anomalien:
 - ▶ **Einfüge-Anomalie**
 - ▶ Eine neue Anzeigenrubrik kann nur dann eingetragen werden, wenn auch eine zugehörige Anzeige aufgenommen wird, denn der Primärschlüssel für den Datensatz muss gesetzt werden
 - ▶ **Lösch-Anomalie**
 - ▶ Wenn Anzeige 507 gelöscht wird, geht die Rubrik Computer verloren
 - ▶ **Änderungs-Anomalie**
 - ▶ Falls sich die E-Mail Adresse von Jens ändert, sind mehrere Datensätze zu ändern



(Relationale) Entwurfstheorie / Ausgangspunkt

- ▶ Welchen prinzipiellen Anforderungen müssen DB-Schemata genügen, damit die eben genannten (und weitere) Anomalien nicht auftreten?
 - ▶ Ursprünglich wurden von Codd (1971/72) drei Normalformen vorgeschlagen (1NF, 2NF, 3NF)
 - ▶ Jede „Stufe“ ($1NF \rightarrow 2NF \rightarrow 3NF$) stellt eine Verschärfung hinsichtlich der Anforderungen (und damit der „Qualität“ des Schemas) bezogen auf die vorherige Stufe dar
 - ▶ Der Normalisierungsprozess führt i. A. zur Aufspaltung (Zerteilung) von Relationen
 - ▶ Definition der 3. Normalform hatte einige Schwächen und wurde später (ca. 1974) revidiert (\rightarrow Boyce-Codd-NF, (BCNF)).
 - ▶ Später kamen viele weitere, spezielle Normalformen hinzu
 - ▶ Fagin führte 1977 eine „neue“ Normalform 4 und 1979 eine „projectionjoin normal form“ 5 ein, die heute überwiegend als 4. und 5. Normalform bezeichnet werden



Normalformen

Gesamtheit der "Relationen" (normalisierte und nicht-normalisierte)

1NF-Relationen

2NF-Relationen

3NF-Relationen

BCNF-Relationen

4NF-Relationen

5NF-Relationen



-
- ▶ Wichtige Fragestellungen darüber hinaus:
 - ▶ Welche Zerlegungen sind verlustfrei ?
 - ▶ kein Verlust von Information
 - ▶ keine „neuen“ (falschen) Informationen
 - ▶ Welche Zerlegungen sind abhängigkeitsbewahrend?
 - ▶ kein Verlust von Integritätsbedingungen
 - ▶ Hinweise:
 - ▶ Normalformen sind Regeln/Hinweise („guidelines“) für guten DB-Entwurf, kein „Kochrezept“
 - ▶ Strikte Normalisierung führt i.d.R. zu einer größeren Anzahl von Relationen
 - ▶ Normalisierung erfordert Wissen über die Bedeutung von Attributen und ihre (konkreten!) Zusammenhänge
 - ▶ Je weiter (2NF → 3NF → ...) normalisiert werden soll, desto höher sind die Anforderungen an das Hintergrundwissen über die Daten
 - ▶ Nach der Normalisierung sollte geprüft werden, ob evtl. Zusammenfassungen möglich sind (→ „Relationen-Synthese“)
-



Einschub: Funktionale Abhängigkeit (functional dependencies, FD's)

► Definition: Funktionale Abhängigkeit (FD)

- $\text{sch}(R)$... Menge der Attribute der Relation R .
- Seien $X, Y \subseteq \text{sch}(R)$. Eine Funktionale Abhängigkeit (FD) $X \rightarrow Y$ ist in R erfüllt, wenn für alle Ausprägungen $\text{val}(R)$ gilt: $t_1, t_2 \in \text{val}(R) : t_1(X) = t_2(X) \Rightarrow t_1(Y) = t_2(Y)$
- In Worten:
- Falls $t_1(X) = t_2(X)$, dann muss auch $t_1(Y) = t_2(Y)$ gelten

► Eine Funktionale Abhängigkeit (FD) wird folgendermaßen geschrieben:

- $X \rightarrow Y$
- Sprechweisen:
- „ X bestimmt Y (funktional)“
- „ Y hängt von X ab“



Funktionale Abhängigkeit: Beispiel

- ▶ Gegeben sind folgende Relation funktionale Abhängigkeiten. Entscheide jeweils, ob in der Relation die angegebene funktionale Abhängigkeit gilt und begründe kurz!

- ▶ $A \rightarrow BC$

- ▶ $AB \rightarrow C$

| A | B | C |
|---|---|---|
| 3 | 7 | 8 |
| 1 | 2 | 6 |
| 5 | 7 | 8 |
| 9 | 8 | 5 |
| 5 | 7 | 9 |
| 3 | 7 | 8 |

$A \rightarrow BC$ ist nicht funktional abhängig, da es zu einem A verschiedene BC gibt:
 $5 \rightarrow 7,8$ $5 \rightarrow 7,9$

$AB \rightarrow C$ ist nicht funktional abhängig, da es zu einem AB verschiedene C gibt:
 $5,7 \rightarrow 8$ $5,7 \rightarrow 9$

Funktionale Abhängigkeit: Übung

- ▶ Gegeben sind folgende Relation funktionale Abhängigkeiten. Entscheide jeweils, ob in der Relation die angegebene funktionale Abhängigkeit gilt und begründe kurz!

- ▶ $AD \rightarrow BC$

- ▶ $B \rightarrow D$

| A | B | C | D |
|---|---|---|---|
| 4 | 5 | 7 | 3 |
| 6 | 9 | 6 | 4 |
| 6 | 8 | 7 | 3 |
| 4 | 6 | 7 | 3 |
| 2 | 7 | 4 | 3 |
| 2 | 8 | 5 | 3 |



Funktionale Abhängigkeit: Übung

- ▶ Gegeben sind folgende Relation funktionale Abhängigkeiten. Entscheide jeweils, ob in der Relation die angegebene funktionale Abhängigkeit gilt und begründe kurz!

- ▶ $AB \rightarrow AB$

- ▶ $C \rightarrow AB$

| A | B | C |
|---|---|---|
| 4 | 4 | 7 |
| 6 | 9 | 6 |
| 6 | 8 | 7 |
| 3 | 2 | 5 |
| 3 | 1 | 5 |



Funktionale Abhängigkeit: Übung

- ▶ $R = (\text{PERSONALNR}, \text{ABTEILUNG}, \text{ADRESSE})$
- ▶ $F = \{\text{PERSONALNR} \rightarrow \text{ABTEILUNG}, \text{ABTEILUNG} \rightarrow \text{ADRESSE}\}$

▶ a)

| PERSONALNR | ABTEILUNG | ADRESSE |
|------------|-------------|-----------|
| 12 | Buchhaltung | 1010 Wien |
| 13 | Verkauf | 1030 Wien |
| 11 | Filiale | 8010 Graz |
| 14 | Buchhaltung | 1010 Wien |

▶ b)

| PERSONALNR | ABTEILUNG | ADRESSE |
|------------|-------------|---------------|
| 12 | Buchhaltung | 1010 Wien |
| 13 | Verkauf | 1030 Wien |
| 11 | Filiale | 8010 Graz |
| 14 | Buchhaltung | 5010 Salzburg |

Funktionale Abhängigkeit: Übung

- Gegeben ist die Relation $r(R)$:

| r | (A | B | C | D | E) |
|---|----|----|----|----|----|
| | a1 | b1 | c1 | d1 | e1 |
| | a1 | b2 | c2 | d2 | e1 |
| | a2 | b1 | c3 | d3 | e1 |
| | a2 | b1 | c4 | d3 | e1 |
| | a3 | b2 | c5 | d1 | e1 |

- Gib an, welche der folgenden Abhängigkeiten r erfüllt.
- a. $A \rightarrow D$
 - b. $AB \rightarrow D$
 - c. $C \rightarrow BDE$
 - d. $E \rightarrow A$
 - e. $A \rightarrow E$
 - f. $A \rightarrow BC$
-



Normalisierung

- ▶ Redundanzen werden mittels standardisierten Vorgangsweisen der Normalisierung Schritt für Schritt aufgehoben



Erste Normalform

- ▶ Die 1. Normalform verlangt, dass sich ein Attribut nicht weiter unterteilen lässt. Außerdem benötigt eine Tabelle einen Primärschlüssel.
- ▶ Definition:
 - ▶ Eine Relation R ist in der ersten Normalform, wenn all ihre Domänen (Attribute) nur elementare Werte enthalten.
oder
 - ▶ Eine Tabelle befindet sich in der ersten Normalform, wenn alle Attribute atomar vorliegen.
 - ▶ Atomar (griechisch: atomos „unteilbar“) heißt, dass sich ein Attribut nicht in weitere Attribute unterteilen lässt.
- ▶ Zwischenfrage: Im Beispiel ist der Primärschlüssel Anzeigennummer und Rubriknummer. Warum?



Ausgangslage

| <u>Anz.Nr.</u> | Anzeigentext, -datum | Ins.Nr. | Nickname, E-Mail | <u>Rub.Nr.</u> | Rubrikbezeichnung |
|----------------|------------------------------------|---------|---------------------------|----------------|-------------------|
| 501 | Mountainbike, 15.03.2016 | 214 | Sissi sissi@jker.de | 2 | Zweiräder |
| 502 | 5 Kinderbücher, 15.03.2016 | 211 | Franki franki@qwert.it | 5 | Bücher |
| 505 | Harry Potter 5, 014.03.2016 | 211 | Franki franki@qwert.it | 5 | Bücher |
| 507 | Buch TCP/IP Grundl., 02.03.2016 | 210 | Jens jens@wert.au | 3 | Computer |
| 507 | Buch TCP/IP Grundl., 02.03.2016 | 210 | Jens jens@wert.au | 5 | Bücher |
| 507 | Buch TCP/IP Grundl., 02.03.2016 | 210 | Jens jens@wert.au | 6 | Sonstiges |
| 508 | Gebr. Trekkingrad, 02.03.2016 | 214 | Sissi sissi@jker.de | 2 | Zweiräder |



Erste Normalform am Beispiel

- ▶ Die Attribute Anzeigentext und -datum sowie Nickname und E-Mail erfüllen nicht die 1. Normalform. Deshalb werden zwei neue Attribute eingeführt: Anzeigendatum und E-Mail



Erste Normalform am Beispiel

| <u>Anz.Nr.</u> | Anzeigentext | Anzeigen- datum | Ins.Nr. | Nickname | E-Mail | <u>Rub. Nr.</u> | Rubrikbe- zeichnung |
|----------------|------------------------|--------------------|---------|----------|-----------------|---------------------|------------------------|
| 501 | Mountainbike | 15.03.2016 | 214 | Sissi | sissi@jker.de | 2 | Zweiräder |
| 502 | 5 Kinderbücher | 15.03.2016 | 211 | Franki | franki@qwert.it | 5 | Bücher |
| 505 | Harry Potter 5 | 14.03.2016 | 211 | Franki | franki@qwert.it | 5 | Bücher |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au | 3 | Computer |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au | 5 | Bücher |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au | 6 | Sonstiges |
| 508 | Gebr. Trekkingrad | 02.03.2016 | 214 | Sissi | sissi@jker.de | 2 | Zweiräder |



Erste Normalform: Übung

- Überprüfe ob folgende Tabelle in der ersten Normalform vorliegt und wandle sie gegebenenfalls um!

Studenten

| Vorname | Nachname | Informatikkenntnisse |
|---------|----------|----------------------|
| Thomas | Müller | Java, C++, PHP |
| Ursula | Meier | PHP, Java |
| Igor | Müller | C++, Java |

Studenten

| Vorname | Nachname | Informatikkenntnisse |
|---------|----------|----------------------|
| Thomas | Müller | C++ |
| Thomas | Müller | PHP |
| Thomas | Müller | Java |
| Ursula | Meier | Java |
| Ursula | Meier | PHP |
| Igor | Müller | Java |
| Igor | Müller | C++ |

Zweite Normalform

► Definition:

- Eine Tabelle befindet sich in der 2. Normalform, wenn die 1. Normalform erfüllt ist und jedes nicht dem Primärschlüssel angehörige Attribut funktional vom Primärschlüssel, aber nicht von Teilen des Primärschlüssels abhängt.
 - oder
- Eine Tabelle befindet sich in der 2. Normalform, wenn die 1. Normalform erfüllt ist und jedes nicht dem Primärschlüssel angehörige Attribut voll funktional vom Primärschlüssel abhängt.

► Definition: funktional abhängig

- Gegeben sei Relation $R(A_1, \dots, A_n)$ und X, Y Teilmengen aus $\{A_1, \dots, A_n\}$:
Eine Attributkombination Y heißt funktional abhängig von X , (in Zeichen $X \rightarrow Y$) wenn in jedem möglichen Tupel von R die Werte von Y durch die von X eindeutig bestimmt sind.

► Definition: voll funktional abhängig

- X, Y wie oben. Eine Attributkombination Y heißt voll funktional abhängig von X , wenn Y funktional abhängig von X ist, aber nicht funktional abhängig von einer echten Teilmenge von X ist (in Zeichen: $X \twoheadrightarrow Y$).
- Funktionale Abhängigkeit liegt vor, wenn zu einem bestimmten Wert eines Attributs höchstens ein Wert des abhängigen Attributs möglich ist.



Beispiel: Ausgangslage nach der ersten Normalform

| <u>Anz.Nr.</u> | Anzeigentext | Anzeigen- datum | Ins. Nr. | Nickname | E-Mail | <u>Rub. Nr.</u> | Rubrikbe- zeichnung |
|----------------|------------------------|--------------------|-------------|----------|-----------------|---------------------|------------------------|
| 501 | Mountainbike | 15.03.2016 | 214 | Sissi | sissi@jker.de | 2 | Zweiräder |
| 502 | 5 Kinderbücher | 15.03.2016 | 211 | Franki | franki@qwert.it | 5 | Bücher |
| 505 | Harry Potter 5 | 14.03.2016 | 211 | Franki | franki@qwert.it | 5 | Bücher |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au | 3 | Computer |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au | 5 | Bücher |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au | 6 | Sonstiges |
| 508 | Gebr. Trekkingrad | 02.03.2016 | 214 | Sissi | sissi@jker.de | 2 | Zweiräder |



Zweite Normalform am Beispiel

- ▶ Die Attribute Anzeigentext, Anzeigendatum, Ins.Nr., Nickname und E-Mail sind funktional vom Attribut Anzeigennummer abhängig. Da in der 2. Normalform eine funktionale Abhängigkeit vom gesamten Primärschlüssel – also Anzeigennummer und Rubriknummer – gefordert ist, ist die 2. Normalform verletzt.
- ▶ Die 2. Normalform lässt sich herstellen, indem die besprochenen Attribute in eine neue Tabelle ausgelagert werden. Im Beispiel bleiben in der Ausgangstabelle nur noch die Attribute Anzeigennummer und Rubriknummer übrig, da kein Attribut der Ausgangstabelle vom gesamten Primärschlüssel abhängig ist.



Zweite Normalform am Beispiel

| <u>Anz.Nr.</u> | <u>Rub.Nr.</u> |
|----------------|----------------|
| 501 | 2 |
| 502 | 5 |
| 505 | 5 |
| 507 | 3 |
| 507 | 5 |
| 507 | 6 |
| 508 | 2 |

| <u>Rub.Nr.</u> | Rubrikbezeichnung |
|----------------|-------------------|
| 2 | Zweiräder |
| 5 | Bücher |
| 3 | Computer |
| 6 | Sonstiges |

| <u>Anz.Nr.</u> | Anzeigentext | Anzeigen- datum | Ins.Nr. | Nickname | E-Mail |
|----------------|---------------------|--------------------|---------|----------|-----------------|
| 501 | Mountainbike | 15.03.2016 | 214 | Sissi | sissi@jker.de |
| 502 | 5 Kinderbücher | 15.03.2016 | 211 | Franki | franki@qwert.it |
| 505 | Harry Potter 5 | 14.03.2016 | 211 | Franki | franki@qwert.it |
| 507 | Buch TCP/IP Grundl. | 02.03.2016 | 210 | Jens | jens@wert.au |
| 508 | Gebr. Trekkingrad | 02.03.2016 | 214 | Sissi | sissi@jker.de |

Zweite Normalform: Beispiel

▶ Beispiel:

- ▶ Proj_Mitarbeiter(Pers#, Proj#, N_name, ..., Proj_name, ...)

▶ Diese Relation ist in 1. NF, aber nicht in 2. NF, denn:

- ▶ Alle Attribute die zur Person gehören, werden durch Pers# eindeutig identifiziert.
- ▶ Alle Attribute die zum Projekt gehören, werden durch Proj# eindeutig identifiziert.

▶ Auflösung:

- ▶ 3 Relationen, sind alle mindestens in 2. NF:

- ▶ Person (Pers#, N_name, ...)
- ▶ Projekt (Proj#, Proj_name, ...)
- ▶ Proj_Mitarbeiter (Pers#, Proj#)



Zweite Normalform: Anmerkungen

- ▶ Relationen mit nur einem Schlüsselattribut sind stets in 2NF.
- ▶ Relationen, die nur aus Schlüsselattributen bestehen, sind ebenfalls stets in 2NF.
- ▶ Relationen, die nicht in 2NF sind,
 - ▶ führen i.d.R. zu redundanter Speicherung von Information (→ Anomalien)
 - ▶ sind Kandidaten für eine Zerlegung



Zweite Normalform: Übung

- ▶ Zur Verwaltung der Prüfungsnoten sei folgende Relation gegeben:
 - ▶ Student(IDSt, StudentNachname, IDProf, ProfessorNachname, Note)
 - ▶ Alle Attribute sind einfach und einwertig.
- ▶ Zudem ist bekannt, dass folgende funktionale Abhängigkeiten existieren:
 - ▶ 1. Das Attribut „ProfessorNachname“ ist funktional abhängig vom Attribut „IDProf“ („IDProf --> ProfessorNachname“)
 - ▶ 2. Das Attribut „StudentNachname“ ist funktional abhängig vom Attribut „IDSt“ („IDSt --> StudentNachname“)
 - ▶ 3. Das Attribut „Note“ ist voll funktional abhängig von den Attributen „IDSt“ und „IDProf“ („IDSt, IDProf ==> Note“)



Zweite Normalform: Übung

- Überprüfe ob die Tabelle in der zweiten Normalform vorliegt und wandle sie gegebenenfalls um!

Studenten

| IDSt | Nachname | IDProf | Professor | Note |
|------|----------|--------|------------|------|
| 1 | Müller | 3 | Schmid | 5 |
| 2 | Meier | 2 | Borner | 4 |
| 3 | Tobler | 1 | Bernasconi | 6 |

