

Wirtschaftsinformatik 2

LE 10 – Normalformen

Prof. Dr. Thomas Off

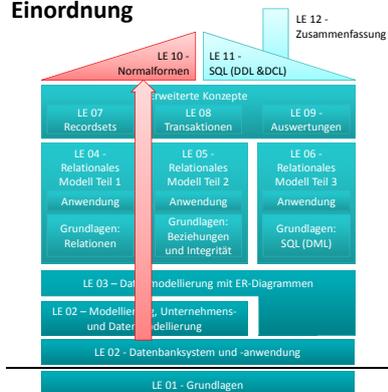
<http://www.ThomasOff.de/lehre/beuth/wi2>

Ziel

Ziel dieser Lehreinheit

- Normalformen und Normalisierung kennenlernen
- Verstehen der Herausforderungen und Probleme die sich beim Entwurf von Datenbanken aus den an sie gestellten Anforderungen ergeben
- Anwenden der Normalisierung auf vorhandene Relationen, um die Anforderungen besser zu erreichen
- Analysieren und Beurteilen eines Datenbankentwurfs hinsichtlich des Grades der Normalisierung

Einordnung



Inhalt

Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Wirtschaftsinformatik 2 - LE 10 - Normalformen 4

Rückblick (LE09 – Auswertungen)

Berichte in MS Access

- werden in aufbereiteter Form als Dokument mit relevanten Daten, mit Gruppierungs- und Summenfunktionen und mit einem Layout erzeugt und zum Export bereitgestellt

Exemplarische Vorgehensweise zur Erstellung von Berichten

- Berichtsstruktur entwerfen
- Erstellung von Abfragen und Teilberichten
 - Erstellen geeigneter Abfragen, die notwendige Angaben liefern
 - Erstellung eines einzelnen Berichtes pro Abfrage
- Zusammenfassung der einzelnen Berichte zu einem Gesamtbericht mit Sortierung, Summen, Gruppierungen
- Einbettung der Berichtserzeugung in den Programmcode (DoCmd.OpenReport)

Wirtschaftsinformatik 2 - LE 10 - Normalformen 5

Rückblick (LE09 – Auswertungen)

SQL bietet Statistikfunktionen, z.B. AVG(), COUNT(), MAX(), MIN(), SUM()

Unterabfragen: Mittel zur Auswertung von Zusammenhängen

```
SELECT COUNT(kndOrt) AS AnzahlWohnorte
FROM
  (SELECT DISTINCT kndOrt
   FROM tblKunden) AS Unterabfrage;
```

Gruppierung von Daten

- GROUP BY: Spalten, anhand derer Tupel mit gleichen Werten gruppiert werden
- HAVING: Filterung anhand von Werten, die sich durch die Gruppenbildung ergeben

```
SELECT kndOrt, COUNT(*) AS KundenImOrt FROM tblKunden
GROUP BY kndOrt
HAVING COUNT(*) > 7;
```

Wirtschaftsinformatik 2 - LE 10 - Normalformen 6

Rückblick



LE 09 - ~~Normalformen~~



Rückblick (LE04 – Relationales Modell Teil 1)



Definition "Schlüsselkandidat"

- besteht aus einem oder mehreren Attributen
- identifiziert eindeutig jedes Tupel (Datensatz)
- ist minimal, d.h. beim Weglassen eines Attributes geht Eindeutigkeit verloren

Definition "Primärschlüssel" (syn. "Primary Key")

- besitzt eine Relation mehr als einen Schlüsselkandidaten, wird einer als Primärschlüssel ausgewählt
- andere Schlüsselkandidaten sind "Alternativschlüssel"

Rückblick



LE 02 - ~~Beziehungen~~



Rückblick (LE 02 – Grundlagen DBS)



Anforderungen an relationale Datenbanken

- vermeiden die Mehrfachspeicherung (Redundanz) gleicher Daten
 - keine unnötige Belegung von Speicherplatz
 - Vermeidung von Inkonsistenzen/Widersprüchen wenn nicht alle Vorkommen gleich und gleichzeitig geändert werden
- einfache und sichere Änderungsmöglichkeiten der Daten
 - Reduzierung von Abhängigkeiten
 - Reduzierung des Risikos von Inkonsistenzen/Widersprüchen
- ...

Inhalt



Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Einführung der Normalisierung



Verschiedene Ansätze

Relationen entwerfen, Bsp.:

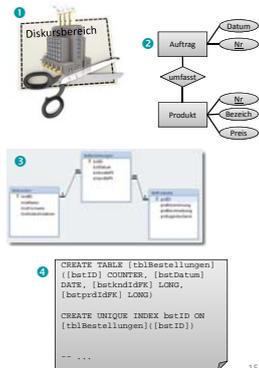
- Strukturierter Ansatz durch Ableitung aus ER-Modellen
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen
- ...



Einführung der Normalisierung

Verschiedene Ansätze Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
 - führt zu mehreren Tabellen, die
 - einzelne Sachverhalte des Diskusbereichs abbilden und
 - in Beziehung stehen.
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen



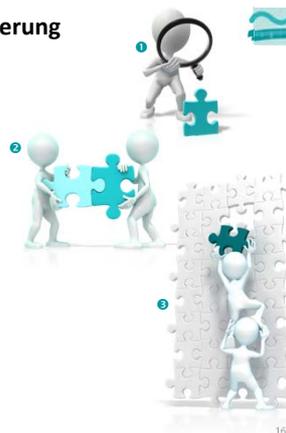
Wirtschaftsinformatik 2 - LE 10 - Normalformen

15

Einführung der Normalisierung

Verschiedene Ansätze Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen
 - führt (initial) zu Tabellen mit vielen Attributen, ohne dass einzelne Sachverhalte oder Beziehungen erkennbar sind



Wirtschaftsinformatik 2 - LE 10 - Normalformen

16

Einführung der Normalisierung

Verschiedene Ansätze Relationen entwerfen, Bsp.:

- Strukturierter Ansatz durch Ableitung aus ER-Modellen
- Sammlung zu speichernder Informationen und deren Zusammenfassung in Tabellen
- ...



Wirtschaftsinformatik 2 - LE 10 - Normalformen

17

Normalformen



Normalformen sind Regeln

- die helfen, Fehler im Entwurfsprozess zu vermeiden
- mit denen die Ziele relationaler Datenbanken erreicht werden können

Normalisierung als Prozess, diese Regeln anzuwenden

- um schrittweise Redundanzen und (nicht gewünschte) Abhängigkeiten aufzuspüren
- der in definierten Stufen (Normalformen) durchlaufen wird und
- in dem Anpassungen der Relationenstruktur vorgenommen werden, ohne dass Informationsverluste entstehen.

Normalformen



Ausgangssituation



Normalformen



Ausgangssituation



"Verkäufer (Nr, Name) verkaufen immer mindestens ein Produkt. Es gibt eine Priorität, in der ein Verkäufer sich um den Verkauf von Produkten (Bezeichnung) kümmert. Der Verkäufer erzielt Umsatz durch den Verkauf von Produkten. Jeder Verkäufer gehört zu genau einer Filiale (Nr, PLZ, Ort)."

Hintergrund



Anomalien in Datenbanken

- im Allgemeinen ein "Unregelmäßigkeit", Abweichung von üblichen Regeln
- Anomalien können in Datenbanken Auslöser von Inkonsistenzen/Problemen sein

Hintergrund



Einfügeanomalie

- Entität kann nicht oder nur schwer zu einer Tabelle hinzugefügt werden,
- weil für den Datensatz noch nicht bekannte Werte (z.B. als Teil des Primärschlüssels) benötigt werden.

Änderungsanomalie

- tritt auf, wenn eine Entität mehrfach gespeichert wurde (Redundanz) und dadurch die mehrfache Änderung gleicher Werte notwendig ist, obwohl sich nur ein Sachverhalt geändert hat und damit das Risiko von Inkonsistenzen steigt.

Löschanomalie

- beim Löschen einer Entität gehen Informationen im zugehörigen Datensatz verloren, die noch benötigt werden und nicht hätten gelöscht werden sollen.

Hintergrund



Anomalien in Datenbanken

- Hinweis: Beispiele für weitere Arten von Anomalien, die durch mehrere Benutzer verursacht werden können, wurden in LE 08 vorgestellt (Lost Update, Dirty Read, Non-Repeatable Read, Phantom Read).

Normalformen

2. Normalform

- Ergebnis der Normalisierung
 - Aufteilung auf zwei Relationen
 - Verkäufer mit Primärschlüssel VKNr, nicht zusammengesetzt
 - Produkte
 - mit zusammengesetztem Primärschlüssel VKNr, Prio
 - VKNr ist gleichzeitig Fremdschlüssel für Zuordnung zu Verkäufer

Verkäufer	VKNr	Name	Filiale	Plz	Ort
123	Albers	1	12345	Berlin	
234	Boehrs	1	12345	Berlin	
345	Dinkel	2	13456	Berlin	
456	Dinkels	2	13456	Berlin	

Produkte	VKNr	Prio	Produkt	Umsatz
123	1	1	Kühlschrank	61.000€
123	2	1	Waschmaschine	59.000€
234	1	1	TV	12.500€
234	2	1	Kühlschrank	55.000€
234	3	1	DVD	10.500€
345	1	1	TV	56.000€
456	1	1	Kühlschrank	12.000€

Wirtschaftsinformatik 2 - LE 10 - Normalformen

48

Normalformen

... ist das ein guter Entwurf? Bezogen auf die Ziele

- komfortable und schnelle Abfragemöglichkeiten
- leichte und einfache Änderbarkeit
- Redundanzfreiheit (um z.B. Speicherplatz zu sparen, Änderungen zu vereinfachen)

Verkäufer	VKNr	Name	Filiale	Plz	Ort
123	Albers	1	12345	Berlin	
234	Boehrs	1	12345	Berlin	
345	Dinkel	2	13456	Berlin	
456	Dinkels	2	13456	Berlin	

Produkte	VKNr	Prio	Produkt	Umsatz
123	1	1	Kühlschrank	61.000€
123	2	1	Waschmaschine	59.000€
234	1	1	TV	12.500€
234	2	1	Kühlschrank	55.000€
234	3	1	DVD	10.500€
345	1	1	TV	56.000€
456	1	1	Kühlschrank	12.000€

Wirtschaftsinformatik 2 - LE 10 - Normalformen

49

Normalformen

... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Umzug einer Filiale (mehrfache Änderung nötig)
- Änderung der Produktbezeichnung (mehrfache Änderungen nötig)

Verkäufer	VKNr	Name	Filiale	Plz	Ort
123	Albers	1	12345	Berlin	
234	Boehrs	1	12345	Berlin	
345	Dinkel	2	13456	Berlin	
456	Dinkels	2	13456	Berlin	

Produkte	VKNr	Prio	Produkt	Umsatz
123	1	1	Kühlschrank	61.000€
123	2	1	Waschmaschine	59.000€
234	1	1	TV	12.500€
234	2	1	Kühlschrank	55.000€
234	3	1	DVD	10.500€
345	1	1	TV	56.000€
456	1	1	Kühlschrank	12.000€

Wirtschaftsinformatik 2 - LE 10 - Normalformen

50

Hintergrund

Determinante

- Attribut **d** ist eine Determinante, wenn ein anderes Attribut der gleichen Relation voll funktional abhängig von **d** ist
- bedeutet vereinfacht, dass alle Attribute, von denen ein Doppelpfeil ausgeht, Determinanten sind
- Hinweis: **d** kann zusammengesetztes Attribut sein

¹⁾ vgl. hier und im Folgenden [1], S. 126 f.
Wirtschaftsinformatik 2 - LE 10 - Normalformen

52

Hintergrund

Beispiel für Determinanten

- Relation Verkäufer
 - VKNr ⇒ Name; VKNr ist Determinante für Name
 - Filiale ⇒ Plz; Filiale ist Determinante für Plz
 - Filiale ⇒ Ort; Filiale ist Determinante für Ort

Negativbeispiel – keine Determinanten

- Relation Verkäufer:
 - Ort ⇒ Filiale; Ort ist keine Determinante für Filiale

						Produkte	VKNr	Prio	Produkt	Umsatz
						123	1	Kühlschrank	61.000€	
						123	2	Waschmaschine	59.000€	
						234	1	TV	12.500€	
						234	2	Kühlschrank	55.000€	
						234	3	DVD	10.500€	
						345	1	TV	56.000€	
						456	1	Kühlschrank	12.000€	

Verkäufer	VKNr	Name	Filiale	Plz	Ort
123	Albers	1	12345	Berlin	
234	Boehrs	1	12345	Berlin	
345	Dinkel	2	13456	Berlin	
456	Dinkels	2	13456	Berlin	

Wirtschaftsinformatik 2 - LE 10 - Normalformen

53

Normalformen

... ist das ein guter Entwurf?

Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Umzug einer Filiale (mehrfache Änderung nötig)
- Änderung der Produktbezeichnung (mehrfache Änderungen nötig)

						Produkte	VKNr	Prio	Produkt	Umsatz
						123	1	Kühlschrank	61.000€	
						123	2	Waschmaschine	59.000€	
						234	1	TV	12.500€	
						234	2	Kühlschrank	55.000€	
						234	3	DVD	10.500€	
						345	1	TV	56.000€	
						456	1	Kühlschrank	12.000€	

Verkäufer	VKNr	Name	Filiale	Plz	Ort
123	Albers	1	12345	Berlin	
234	Boehrs	1	12345	Berlin	
345	Dinkel	2	13456	Berlin	
456	Dinkels	2	13456	Berlin	

Wirtschaftsinformatik 2 - LE 10 - Normalformen

54

Normalformen

3. Normalform

- Ergebnis der Normalisierung
 - Weitere Relation Filiale
 - Verkäufer mit Fremdschlüssel für Filiale

Filiale	FNr	Plz	Ort
1	12345	Berlin	
2	13456	Berlin	

Verkäufer	VkNr	Name	FNr
123	Albers	1	
234	Boehrs	1	
345	Dinkel	2	
456	Dinkels	2	

Produkte	VkNr	Prio	Produkt	Umsatz
123	1	Kühlschrank	61.000€	
123	2	Waschmaschine	59.000€	
234	1	TV	12.500€	
234	2	Kühlschrank	55.000€	
234	3	DVD	10.500€	
345	1	TV	56.000€	
456	1	Kühlschrank	12.000€	

Wirtschaftsinformatik 2 - LE 10 - Normalformen

58

Normalformen

... ist das ein guter Entwurf? Bezogen auf die Ziele

- komfortable und schnelle Abfragemöglichkeiten
- leichte und einfache Änderbarkeit
- Redundanzfreiheit (um z.B. Speicherplatz zu sparen, Änderungen zu vereinfachen)

Filiale	FNr	Plz	Ort
1	12345	Berlin	
2	13456	Berlin	

Verkäufer	VkNr	Name	FNr
123	Albers	1	
234	Boehrs	1	
345	Dinkel	2	
456	Dinkels	2	

Produkte	VkNr	Prio	Produkt	Umsatz
123	1	Kühlschrank	61.000€	
123	2	Waschmaschine	59.000€	
234	1	TV	12.500€	
234	2	Kühlschrank	55.000€	
234	3	DVD	10.500€	
345	1	TV	56.000€	
456	1	Kühlschrank	12.000€	

Wirtschaftsinformatik 2 - LE 10 - Normalformen

59

Normalformen

... ist das ein guter Entwurf? Fortschritte erkennbar, aber Redundanzen erschweren z.B.:

- Änderung der Produktbezeichnung (mehrfache Änderung nötig)

Filiale	FNr	Plz	Ort
1	12345	Berlin	
2	13456	Berlin	

Verkäufer	VkNr	Name	FNr
123	Albers	1	
234	Boehrs	1	
345	Dinkel	2	
456	Dinkels	2	

Produkte	VkNr	Prio	Produkt	Umsatz
123	1	Kühlschrank	61.000€	
123	2	Waschmaschine	59.000€	
234	1	TV	12.500€	
234	2	Kühlschrank	55.000€	
234	3	DVD	10.500€	
345	1	TV	56.000€	
456	1	Kühlschrank	12.000€	

Wirtschaftsinformatik 2 - LE 10 - Normalformen

61

Normalformen

... ist das ein guter Entwurf?
Ja.



Filiale	FNr	Plz	Ort
1	12345	Berlin	
2	13456	Berlin	

Produkte	PNr	Produkt
1		Kühlschrank
2		Waschmaschine
3		TV
4		DVD

Verkäufer	VkNr	Name	FNr
123	Albers	1	
234	Boehrs	1	
345	Dinkel	2	
456	Dinkels	2	

Umsätze	VkNr	Prio	PNr	Umsatz
123	1	1		61.000€
123	2	2		59.000€
234	1	3		12.500€
234	2	1		55.000€
234	3	4		10.500€
345	1	3		56.000€
456	1	1		12.000€

Wirtschaftsinformatik 2 - LE 10 - Normalformen 62

Normalformen

Weiterführende Normalformen

- 4. Normalform
- 5. Normalform

sind in diesem Kurs nicht relevant.

Wirtschaftsinformatik 2 - LE 10 - Normalformen 63

Hinweise

Häufig wird die Frage nach der Abhängigkeit von PLZ und Ort gestellt...

- Es gibt PLZ, zu denen genau ein Ort gehört (14612 Falkensee)
- Es gibt PLZ, zu denen mehrere Orte gehören (14532)
- Es gibt Orte, zu denen mehrere PLZ gehören (Berlin)

... deshalb gibt es keine funktionale Abhängigkeit des Ortes von der PLZ.

Gebiet	Plz	Ort
10115	Berlin	
14119	Berlin	
14532	Stahnsdorf	
14532	Kleinmachnow	
14612	Falkensee	



© OpenStreetMap, Wikimedia

Wirtschaftsinformatik 2 - LE 10 - Normalformen 64

Hinweise

Ist die Tabelle Filiale wirklich in 3. Normalform?

- Zu jeder Filialnummer existiert höchstens eine Wertekombination aus PLZ und Ort, d.h. es liegt funktionale Abhängigkeit vor
- Filialnummer ist nicht zusammengesetzt, d.h. es liegt volle funktionale Abhängigkeit vor
- Es gibt keine funktionale Abhängigkeit des Ortes von der PLZ, d.h. FNR ist die einzige Determinante

Filiale	FNR	Plz	Ort
1	10115	Berlin	Berlin
2	10115	Berlin	Berlin
3	14476	Potsdam	Potsdam
4	14482	Potsdam	Potsdam
5	14552	Michendorf	Michendorf
6	14552	Saarmund	Saarmund
7	14552	Fresdorf	Fresdorf

Verschiedene Filiale in verschiedenen PLZ-Gebieten des gleichen Ortes

Verschiedene Filialen in einem PLZ-Gebiet des gleichen Ortes.

Verschiedene Filialen in einem PLZ-Gebiet verschiedener Ortes.

Hinweise

Funktionale Abhängigkeit

- kann anhand von Beispieldaten in einer Tabelle nur widerlegt werden
- kann nur anhand des Wissens über den Anwendungsbereich abgeleitet werden

Normalisierung unabhängig vom Ansatz

Exemplarischer Ausgangspunkt hier

- Sammlung zu speichernder Informationen und
- deren Zusammenfassung in Tabellen (Universalrelationen)

... und dennoch

- Normalisierung unabhängig vom gewählten Ansatz und auch bei strukturierter Ableitung aus ER-Modellen sinnvoll



Normalisierung abhängig vom Einsatzzweck

Normalisierung

- Ziel: Beseitigung von Anomalien (bei Einfügen, Löschen und Ändern) insbesondere durch Redundanzfreiheit
- führt zu Redundanzfreiheit durch Relationen mit wenigen Attributen
- Auswertung erfordert JOIN-Operationen über verschiedene normalisierte Relationen

Denormalisierung

- Umkehr der Normalisierung und Zusammenfassung von mehreren Relationen in einer gemeinsamen Relation
- sinnvoll zur Leistungsoptimierung, wenn Einfüge-, Lösch- und Änderungsoperationen selten
- Unterstützt insbesondere die Auswertung von Datenbeständen
- Beispiel: Data Warehouse
 - Faktentabellen mit Redundanzen
 - Dimensionstabellen denormalisiert als Sternschema



Wirtschaftsinformatik 2 - LE 10 - Normalformen

69

Inhalt

Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Wirtschaftsinformatik 2 - LE 10 - Normalformen

70

Zusammenfassung

Normalisierung als Prozess der anhand definierter Regeln, der Normalformen der verlustfreien Optimierung dient

Normalformen 1 bis 3¹⁾

- In Anlehnung an:
"Die Wahrheit, die ganze Wahrheit und nichts als die Wahrheit.
So wahr mir Gott helfe!"
- Bedeutet Normalisierung:
"Der Schlüssel, der ganze Schlüssel und nichts als der Schlüssel.
So wahr mir Codd helfe!"
 - alle atomaren Attribute beziehen sich auf den Schlüssel (1. NF)
 - sie beziehen sich auf den gesamten Schlüssel (2. NF)
 - die Attribute hängen nur vom Schlüssel ab und nicht von weiteren Attributen (3. NF)

1) nach Wikipedia: Normalisierung (Datenbanken), http://de.wikipedia.org/wiki/Normalisierung_%28Datenbank%29

Wirtschaftsinformatik 2 - LE 10 - Normalformen

71

Zusammenfassung



Wichtig

- Relation mit nicht zusammengesetztem Primärschlüssel, die sich in 1. NF befindet, befindet sich auch in 2. NF
- Relation, die sich nicht in 2. NF befindet, hat einen zusammengesetzten Primärschlüssel

Antworten auf Frage: "In welcher NF ist Relation R?"

- Ausgehend von 1. NF prüfen
- Ist sie in 1. NF, dann weiter prüfen ob 2. NF
- Ist sie in 2. NF, dann weiter prüfen, ob 3. NF
- ...
- Abschließende Argumentation: Sie ist in n. NF, weil Sie ist aber nicht in n+1. NF, weil

Inhalt



Ziel und Einordnung

Rückblicke

Einführung

Normalformen

- Ausgangssituation
- Erste Normalform
- Zweite Normalform und ihr formaler Hintergrund
- Dritte Normalform und ihr formaler Hintergrund
- Weitere Normalformen

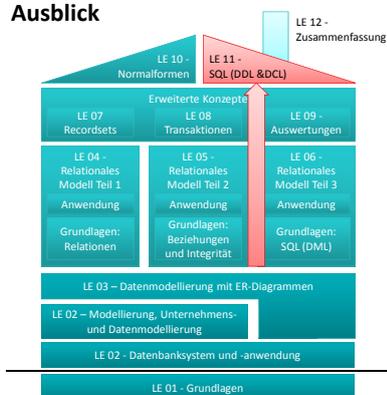
Normalisierung

- Gewählter Ansatz
- Geplanter Einsatzzweck

Zusammenfassung

Ausblick

Ausblick



Quellen



- [1] E. Schicker, E.: Datenbanken und SQL. Teubner, Stuttgart, 1996.
- [2] A. Fink, G. Schneiderreit, S. Voß: Grundlagen der Wirtschaftsinformatik. Physika (Springer), Heidelberg, 2001.
- [3] H. Sauer: Relationale Datenbanken. 5. Aufl., Addison-Wesley, München, 2002.
- [4] G. Vossen: Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme. 5. Aufl., Oldenbourg, München, 2008.
- [5] Wikipedia: Normalisierung (Datenbanken).
http://de.wikipedia.org/wiki/Normalisierung_%28Datenbank%29#Merksspruch

Wirtschaftsinformatik 2 - LE 10 - Normalformen

75



BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Wirtschaftsinformatik 2 LE 10 – Normalformen

Prof. Dr. Thomas Off

<http://www.ThomasOff.de/lehre>
