

Introduction to Data and Knowledge Engineering

3. Übung

Funktionale Abhängigkeiten und Normalformen



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Bemerkungen zu Normalformen

1NF

1NF: alle Attribute sind atomar.

Bemerkungen:

- Nur teilweise formal überprüfbar: Solange Attribut-Typ nicht Bit ist, kann man immer Komponenten sehen. Klar ist, dass z.B. Listen nicht atomar sind.
- Aus dem Konzept atomare Werte entspringt sowohl die Stärke als auch die Schwäche von EDV-Systemen.
- 1NF liegt etwas quer zu anderen Normalformen: z.B wäre 3NF ohne 1NF denkbar, 3NF ohne 2NF eher nicht.

Bemerkungen zu Normalformen

2NF

2NF: 1NF + kein Extra-Attribut hängt nur von einem Teil der Schlüssel-Attribute ab.

Ein *Extra-Attribut* ist ein Attribut, das nicht Teil des Primär-Schlüssels oder eines anderen Schlüsselkandidaten ist.

Bemerkungen:

- Alternative Formulierung für 2NF: Alle Extra-Attribute sind vollständig vom Primär-Schlüssel abhängig.
- B ist *vollständig abhängig* von $A_1..A_n$, wenn $A_1..A_n \rightarrow B$ gilt, dies aber nicht mehr gilt, wenn eines der Attribute A_i aus der linken Seite entfernt wird.
- Der Ausdruck 'Extra-Attribut' ist nicht üblich, stattdessen wird eher 'Nicht-Schlüssel-Attribut' o.ä. gesagt.

Bemerkungen zu Normalformen

3NF

3NF: 2NF + kein Extra-Attribut ist von anderen Extra-Attributen abhängig.

Bemerkungen:

- Standard-Formulierung: Extra-Attribute sind nicht transitive vom Schlüssel abhängig.

C ist *transitiv abhängig* von $A_1..A_n$ wenn es Attribute $B_1..B_n$ gibt, so dass $A_1..A_n \rightarrow B_1..B_n$ und $B_1..B_n \rightarrow C$ gilt.

Da alle Attribute vom Schlüssel abhängen, ist die einfachere Formulierung oben gleichbedeutend.

- Variante von *transitiv abhängig*: Attribut B statt Attribute $B_1..B_n$. Das ist nicht gleichbedeutend, d.h. damit erhält man eine andere 3NF.

Bemerkungen zu Normalformen

BCNF

BCNF: 3NF + ist ein Attribute B vollständig von $A_1..A_n$ abhängig, dann ist $A_1..A_n$ ein Schlüsselkandidat.

Bemerkungen:

- unter BCNF kann es nicht passieren, dass ein Schlüssel-Attribut von anderen Schlüssel-Attributen abhängig ist,
- und auch nicht, dass ein Attribut von einer Menge aus Schlüssel- und Extra-Attributen abhängt

Bemerkungen zu Normalformen

Allgemein

Allgemeine Bemerkung:

- Manche alternative NF-Definition beziehen sich auf die FD-Menge und die Aufteilung der Attribute in Primär-Schlüssel-Attribute und Extra-Attribute. Andere Definition beziehen sich nur auf die FD-Menge.

Das ist ein Grund, warum man häufig verschiedene Varianten der NFs findet.

Mehrwertige Abhängigkeiten $A \twoheadrightarrow B$:

- Jedem Wert für A kann eine Menge von Werten für B zugeordnet werden, welche für jeden Wert der anderen Attribute alle angenommen werden.

- Beispiel:
In einer Schule, in der jeden Tag alle Fächer drankommen (8-20 Uhr), ist die Menge der zwei unbeliebtesten Fächer (B) einer Schülerin (A) unabhängig vom Tag.

N	A	Tag	B
3	Eva	Dienstag	Häckeln
4	Eva	Freitag	Singen
5	Eva	Dienstag	Singen
6	Eva	Freitag	Häckeln
7	Karl	Dienstag	Mathe
8	Karl	Dienstag	Turnen
8	Karl	Donnerstag	Mathe
9	Karl	Donnerstag	Turnen

Aufgabe 3.1

FDs aus Beschreibung Gewinnen

Beschreibung der Welt:

- Ein Professor kann mehrere Vorlesungen halten und eine Vorlesung kann von mehreren Professoren gehalten werden.
- Es gibt zu jeder Vorlesung ein Foliensatz, unabhängig davon, welcher Professor sie hält. Verschiedene Vorlesungen verwenden nicht den gleichen Foliensatz. Jeder Foliensatz hat einen Professor als Autor.
- Die Studenten bekommen zur Vorlesung eine Note.

Aufgabe 3.1

FDs aus Beschreibung Gewinnen

Attribute:

- P: Professor
- V: Vorlesung
- F: Foliensatz
- S: Studenten
- N: Noten

Schlüsselloses Schema $U = (P, V, F, S, N)$
(statt Schema wird auch 'Relation' gesagt. Geschmackssache.)

Erster Punkt ergibt keine FDs.

Zweiter Punkt: $V \rightarrow F$, $F \rightarrow V$, $F \rightarrow P$,

Dritter Punkt: $VS \rightarrow N$

Aufgabe 3.1

FDs aus Beschreibung Gewinnen

Lassen wir S und N beiseite und betrachten nochmal die FDs auf den verbleibenden Attributen V, F und P. Das sind $V \rightarrow F$, $F \rightarrow P$ und $F \rightarrow V$.

Aus $F \rightarrow P$ und $F \rightarrow V$ folgt $F \rightarrow PV$.

D.h., wenn man die Folien kennt, dann kennt man nicht nur die Vorlesung, sondern weiß auch welcher Professor sie hält?

Das kann nicht sein, da es laut Text zu einer Vorlesung mehrere vortragende Professoren geben kann. Komisch.

Die Rolle des Professors in $F \rightarrow P$ ist die des Autors der Folien, oben haben wir in $F \rightarrow PV$ das P aber als Vortragenden verstanden.

Wir müssten ein weiteres Attribut D wie Dozent einführen, wenn diese Rolle auch behandelt werden soll. Die saubere Lösung ist, drei Attribute P wie Professor, D wie Dozent und A wie Autor zu verwenden. Die FDs dazu sind dann $F \rightarrow A$, $D \rightarrow P$ und $A \rightarrow P$. Man erhält auch wieder $F \rightarrow VA$ und damit $F \rightarrow VP$ aber eben nicht $F \rightarrow AD$.

Aufgabe 3.2

FDs aus Daten Ableiten

Die Interpretation des Schema leider verlorengegangen und die Namen der Attribute helfen auf nicht weiter.

Dingsbums

A	B	C	D
a	1	q	r
m	1	b	8
a	1	l	r
b	3	m	0
b	2	p	0
c	3	v	r

Welche FDs ergeben sich aus den vorhandenen Entities?

Aufgabe 3.2

FDs aus Daten Ableiten

In C sind alle Werte verschieden. Also hängen alle anderen Attribute von C ab:

$C \rightarrow ABD$ (und damit auch $C \rightarrow AB \dots CA \rightarrow ABD$ etc.)

C ist auf der linken Seite erledigt. Was ist mit A?

$A \rightarrow D$

Und umgekehrt? Nein, $D \rightarrow A$ gilt nicht. $B \rightarrow X$ geht auch nicht. Was ist mit zwei Attributen auf der linken Seite?

$BD \rightarrow A$

Alle anderen gewinnt man durch Anwendung der Hüllen-Regeln aus diesen drei.

Aufgabe 3.3

FD-Äquivalenzklassen

Bestimmen Sie die Äquivalenzklassen für die FD-Menge F

$$F = \{B \rightarrow C, A \rightarrow DE, B \rightarrow F, F \rightarrow GH, D \rightarrow I\}$$

Sei $[X_1..X_n]$ die Menge der Attribute die äquivalent zu $X_1..X_n$ sind.

$$[B] = \{B\}, [A] = \{A\}, [F] = \{F\}, [D] = \{D\}$$

Seien $[[X_1..X_n \rightarrow Y_1..Y_m]]$ die FDs, deren linke Seite äquivalent zu $X_1..X_n$ sind.

Äquivalenzklassen	Teilschemata	Schlüsselkandidaten
$[[B \rightarrow C]] = \{B \rightarrow C, B \rightarrow F\}$	(BCF)	{B}
$[[A \rightarrow DE]] = \{A \rightarrow DE\}$	(ADE)	{A}
$[[F \rightarrow GH]] = \{F \rightarrow GH\}$	(FGH)	{F}
$[[D \rightarrow I]] = \{D \rightarrow I\}$	(DI)	{D}