

**Gunter Saake
Kai-Uwe Sattler
Andreas Heuer**

Datenbanken

Konzepte und Sprachen

Sechste Auflage

Dieses pdf-Kapitel ist eine kostenlose Ergänzung zum oben genannten Buch, das 2018 bei MITP erschienen ist.

Bitte beachten: Verweise auf Umgebungen B-X beziehen sich auf Teile innerhalb dieses Download-Kapitels. Verweise, die mit Kapitelnummern beginnen, wie 11-31, beziehen sich auf Umgebungen innerhalb des obigen Lehrbuches. Verweise auf Seiten vor 738 beziehen sich ebenfalls auf das zugrundeliegende Lehrbuch.

Literaturhinweise in diesen pdf-Kapiteln beziehen sich auf ein erweitertes Literaturverzeichnis zum Lehrbuch, das auch als kostenlose Ergänzung an derselben Stelle zum Download bereitsteht.



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Bei der Herstellung des Werkes haben wir uns zukunftsbewusst für umweltverträgliche und wiederverwertbare Materialien entschieden. Der Inhalt ist auf elementar chlorfreiem Papier gedruckt.

ISBN 978-3-95845-776-8
6. Auflage 2018

www.mitp.de
E-Mail: mitp-verlag@sigloch.de

Telefon: +49 7953/7189-079
Telefax: +49 7953/7189-082

© 2018 mitp Verlags GmbH & Co. KG, Frechen

Dieses Werk, einschließlich aller seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Lektorat: Sabine Janatschek, Ernst-Heinrich Profener
Sprachkorrektur: Jürgen Dubau, Astrid Langen
Covergestaltung: Christian Kalkert, www.kalkert.de
Bildnachweis Cover: [iStock.com/Shawn Hempel](https://www.iStock.com/ShawnHempel) | [fotolia.com/karpenko_ilia](https://www.fotolia.com/karpenko_ilia)
Satz: Gunter Saake, Magdeburg; Kai-Uwe Sattler, Ilmenau; Andreas Heuer, Rostock
Druck: Westermann Druck Zwickau GmbH

B.3 Erweiterte Modelle und Anfragealgebren

Das relationale Datenmodell hat eine einfache, klare formale Grundlage, aber auch beschränkte Ausdrucksfähigkeiten in der Datenstrukturierung. Insbesondere hierarchisch zusammengesetzte Strukturen müssen „flach“ realisiert werden. In diesem Abschnitt diskutieren wir Erweiterungen des Relationenmodells um strukturierte Attributwerte in sogenannten geschachtelten Relationen sowie semantische Datenmodelle, die weitere Konzepte zu diesen Ansätzen hinzufügen.

B.3.1 Geschachtelte Relationen: Das NF^2 -Modell

Im relationalen Datenmodell müssen alle Daten in Relationen mit unstrukturierten Attributwerten vorliegen. Während der Diskussion des ER-Modells und seiner Erweiterungen hatten wir bereits Beispiele präsentiert, in denen komplexe Attributwerte modelliert wurden. Insbesondere tupel- und mengenwertige Attributwerte werden sogar für das einfache ER-Modell als naheliegende Erweiterung angesehen. Derartige strukturierte Attributwerte müssen im Relationenmodell simuliert werden: Alle Relationen liegen in der ersten Normalform (1NF) vor; Attributwerte sind atomar und von der Relationenalgebra „nicht weiter zerlegbar“.

Eine Behebung dieser Einschränkung ist die Erweiterung auf das Modell der *geschachtelten Relationen*. Im Englischen werden geschachtelte Relationen als *Nested Relations* oder als *NF^2 Relations* bezeichnet. NF^2 steht hierbei für *Non First Normal Form* (NFNF), also nicht in der ersten Normalform vorliegende Relationen. Wir bezeichnen geschachtelte Relationen ebenfalls kurz als *NF^2 -Relationen*.

NF^2 -Relationen erlauben komplexe Attributwerte in dem Sinne, dass Attribute selbst wieder Relationen sein können. Ein Beispiel für eine NF^2 -Relation wird in Abbildung B.42 gezeigt. Das Attribut Weingüter enthält hier jeweils einige Weingüter zur jeweiligen Anbauregion als Untertabelle. Die Weingüter-Daten umfassen wiederum eine Untertabelle mit Weinen, die sie erzeugen.

Region	Weingüter			
	Weingut	Anbaugebiet	Weine	
			WeinID	Name
Kalifornien	Helena	Napa Valley	3456	Zinfandel
			3478	Pinot Noir
	Bighorn	Napa Valley	4961	Chardonnay
Bordeaux	Château La Rose	Saint-Émilion	1042	La Rose Grand Cru
	Château La Pointe	Pomerol		

Abbildung B.42: Beispiel für eine NF^2 -Relation

Ein geschachteltes Relationenschema kann als naheliegende Erweiterung des „flachen“ Falls als Menge von Attributen

$$R = \{A_1, \dots, A_n\}$$

definiert werden. Diese Attribute sind nun einerseits flache Attribute von einem Standarddatentyp wie im normalen Relationenmodell, also $\text{dom}(A_i) \in \mathcal{D}$, wobei \mathcal{D} die Wertebereiche zu den Standarddatentypen sind. Andererseits können sie aber im Gegensatz zum normalen Relationenmodell selbst wieder (rekursiv) durch eine Menge von Attributen bestimmt sein, für ein Attribut A_i kann also gelten:

$$A_i = \{A_{i_1}, \dots, A_{i_m}\}$$

Auch die restlichen Definitionen des Relationenmodells können auf den geschachtelten Fall übertragen werden.

Schließlich muss auch die Relationenalgebra um neue Operationen erweitert werden. In Abschnitt B.3.3.1 wird eine NF^2 -Algebra vorgestellt. Hier sei nur erwähnt, dass insbesondere zwei zusätzliche Operationen zum Schachteln ν (Nestung) und Entschachteln μ (Entnestung) von Relationen benötigt werden.

Ein Beispiel für einen Prototyp, der NF^2 -Relationen realisiert, ist das an der Universität Darmstadt (später ETH Zürich) in der Gruppe von Schek entwickelte DASDBS-System [SPSW90].

B.3.2 PNF-Relationen

Die beliebige Schachtelung von Relationen kann zu unübersichtlichen und fehlerträchtigen Relationen führen, in denen Mengen von Mengen simuliert werden, die keine direkte Entsprechung im flachen Relationenmodell finden. Eine wichtige Teilklasse sind deshalb die sogenannten *PNF-Relationen*. PNF-Relationen können immer entschachtelt durch eine äquivalente 1NF-Relation dargestellt werden.

PNF steht für *Partitioned Normal Form*. Relationen in PNF haben auf jeder Stufe der Schachtelung einen flachen Schlüssel. Abbildung B.43 zeigt links eine geschachtelte Relation in PNF, bei der auf der obersten Ebene das Attribut A Schlüssel ist. Die rechte Relation ist nicht in PNF – auf der obersten Ebene gibt es gar keine atomaren Attribute, und damit erst recht keine flachen Schlüssel.

Abbildung B.44 zeigt die „flache“ Realisierung der PNF-Relation aus Abbildung B.43.

B.3.3 Verallgemeinerte geschachtelte Relationen

Das Modell der geschachtelten Relationen erlaubt mehrere mögliche Verallgemeinerungen. In NF^2 -Relationen ist der Basiskonstruktor die *Relation*, die in

PNF-Relation:	A	D	
		B	C
1	2	3	
	4	2	
2	1	1	
	4	1	
3	1	1	

Nicht PNF:	A	C
	B	D
1	2	
2	3	
1	3	
2	4	

Abbildung B.43: Beispiel für geschachtelte Relation in PNF und Gegenbeispiel

Entnestete Fassung der PNF-Relation:	A	B	C
	1	2	3
	1	4	2
	2	1	1
	2	4	1
	3	1	1

Abbildung B.44: Flache Darstellung der geschachtelten PNF-Relation aus Abbildung B.43

üblichen Typsystemen dem Konstrukt **of tuple of ...** entspricht. Die möglichen Typkonstruktoren können neben den Konstruktoren **set of** und **of** um Listen mittels **of**, Multimengen mittels **of** oder Arrays mittels **of** ergänzt werden. Diese Typkonstruktoren können beliebig kombiniert werden, so dass beispielsweise **set of bag of integer** erlaubt ist.

Das resultierende Datenmodell wird als *erweitertes NF²-Modell* bezeichnet, kurz *eNF²-Modell*. Dieses Modell wird etwa von dem am Wissenschaftlichen Zentrum Heidelberg der IBM entwickelten Prototypen AIM-P realisiert [PT86, DL89, PD89].

Bei den bisherigen Erweiterungen müssen nicht hierarchische Beziehungen über Wertevergleiche wie im flachen Relationenmodell realisiert werden. Eine weitere mögliche Erweiterung ist ein abstrakter Datentyp, der Referenzen auf Tupel aufnehmen kann. Mit dieser Erweiterung kann das formale Modell geschachtelter Relationen um Referenzen erweitert werden, die verzeigerte Strukturen möglich machen. Als Konsequenz entsprechen Attribute allgemeinen Funktionen, die auch „Objekte“ als Werte annehmen können.

Abschliessend wollen wir die Typkonstruktoren zu den bisher behandelten formalen Datenmodellen in Verbindung setzen. Einige dieser Datenmodelle definieren sich über die Kombinierbarkeit der Konstruktoren zur Definition von Datenbankobjekten. Abbildung B.45 verdeutlicht dies, wobei **atomic** für die Basisdatentypen steht.

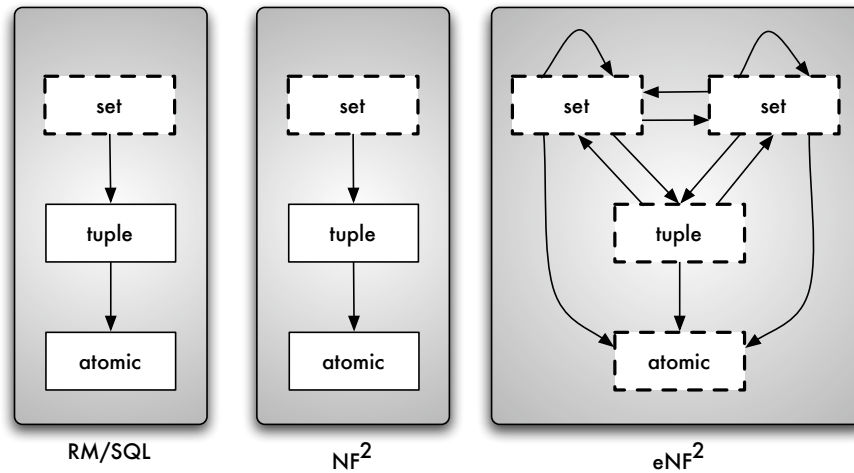


Abbildung B.45: Kombinierbarkeit der Konstruktoren

Das Relationenmodell erlaubt nur Mengen von Tupeln als Datenbankobjekte. SQL schließt sich dem an, erlaubt aber zumindest als Ergebnisse von Operationen auch Multimengen. Geschachtelte Relationen erlauben Relationen, also Mengen von Tupeln, als Attributwerte. Das eNF²-Modell nimmt Listen hinzu. Letzteres Modell erlaubt im Gegensatz zu den ersten beiden jede Kombination als Typ eines im Datenbankkatalogs eingetragenen Datenbankobjekts (angedeutet durch die gestrichelt umrandeten Kästchen).

B.3.3.1 Erweiterte Anfragealgebren

Für erweiterte Anfragekonzepte, insbesondere für objektrelationale Datenbanken und Data-Warehouse-Anwendungen, benötigen wir Erweiterungen der Relationenalgebra um weitere Operatoren bzw. Anpassungen an ein verfeinertes Datenmodell.

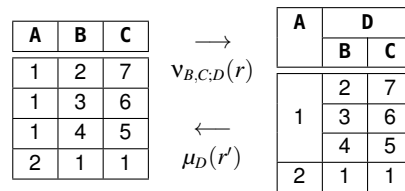
NF²-Algebra

Als Erweiterung des Relationenmodells haben wir bereits in Abschnitt B.3 das NF²-Modell kennengelernt. Passend zu diesem Modell gibt es nun mehrere Vorschläge für Algebren, deren Operationen direkt aus der Relationenalgebra übernommen oder Erweiterungen der Relationenalgebra-Operationen sind:

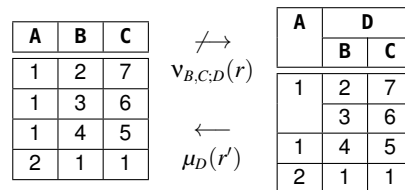
- $\cup, -, \pi, \bowtie$ werden zunächst wie in der Relationenalgebra eingeführt.
- Die σ -Bedingungen werden erweitert um:

- Relationen als Operanden (statt nur Konstanten von Standarddatentypen)
- Mengenvergleiche, wie etwa $\theta: =, \subseteq, \subset, \supset, \supseteq$
- Es gibt jetzt rekursiv aufgebaute Operationsparameter. So können etwa π und σ auch innerhalb von Projektionslisten und Selektionsbedingungen dort angewendet werden, wo relationenwertige Attribute auftauchen.
- Zwei zusätzliche Operationen ν (Nestung, engl. *nest*) und μ (Entnestung, engl. *unnest*) werden eingeführt.

◀**Beispiel B-4**▶ Ein kleines Beispiel soll die neuen Operationen veranschaulichen. Wenden wir auf die unten links stehende Relation die dort angegebene Nestung an, so kommt die NF²-Relation rechts heraus.



In diesem speziellen Fall kann die Nestung durch die Entnestung wieder rückgängig gemacht werden. Leider ist die Nestung nicht allgemein die Inverse der Entnestung. Dies kann man an einer Abwandlung des Beispiel sehen:



Wie man sieht, ist die Inversität insbesondere dann nicht gegeben, wenn die geschachtelte Relation nicht in PNF ist. □

Minimale geschachtelte Algebra

Für eine minimale geschachtelte Algebra werden die Operationen Projektion, Selektion (auch mit Bedingungen auf Mengen), Verbund, Mengenoperationen und Umbenennung aus der Relationenalgebra benötigt. Zusätzlich werden Nestung und Entnestung definiert:

- Die *Nestung* $\nu[(A_1, \dots, A_n); A](r(R))$ fasst Attribute A_1, \dots, A_n des Relationenschemas R zu einem neuen Attribut A zusammen, d.h. A ist definiert als:

$$\text{set of}(\text{tuple of}(A_1, \dots, A_n))$$

Mehrere (A_1, \dots, A_n) -Tupel werden zu einer Menge zusammengefasst, wenn die Werte der Tupel in der Relation r auf den restlichen Attributen des Relationenschemas (also auf $R - \{A_1, \dots, A_n\}$) übereinstimmen.

- Die *Entnestung* $\mu[A](r(R))$ löst ein geschachteltes Attribut A auf, d.h. falls A als

$$\text{set of}(\text{tuple of}(A_1, \dots, A_n))$$

definiert ist, sind im Ergebnis die Attribute A_1, \dots, A_n im Relationenschema enthalten. Die einzelnen Tupel der Attributwerte von A werden zusammen mit den zugehörigen Attributwerten der restlichen Attribute von R zu neuen Tupeln verbunden.

Die Entnestung macht eine Nestung rückgängig; die Umkehrung gilt allerdings nicht immer, wie in Beispiel B-4 zu sehen war.

Die vorgestellte minimale Algebra erfordert ein umständliches Arbeiten: Oft muss erst eine Entnestung vorgenommen werden, bevor andere Operationen ausgeführt werden können und abschließend wieder eine Nestung die ursprüngliche Struktur der Tabelle wiederherstellt. Dieses Manko wird mit der im Folgenden beschriebenen Erweiterung beseitigt.

Orthogonale geschachtelte Algebra

Ein Beispiel für eine erweiterte Algebra ist die Algebra von Schek und Scholl, die in [SS86, SS89] vorgestellt wurde. Die Operationen Projektion und Selektion können nun rekursiv geschachtelt eingesetzt werden. Wir diskutieren diese Möglichkeiten anhand der geschachtelten Relation r in Abbildung B.42 auf Seite 791.

◀**Beispiel B-5**▶ Im ersten Beispiel werden gleichzeitig auf mehreren Ebenen Projektionen durchgeführt:

$$\pi[\text{Region}, \pi[\text{Anbaugebiet}](\text{Weingüter})](r)$$

Das zweite Beispiel zeigt eine Selektion auf einer Unterrelation, die in die Projektion integriert wird:

$$\pi[\text{Region}, \sigma[\text{Anbaugebiet} = \text{Pomerol}](\text{Weingüter})](r)$$

□

In dieser erweiterten Algebra sind Projektion und Selektion *orthogonal* einsetzbar, sofern die Tabellenstruktur es zulässt: Jede beliebige Kombinati-

on (Projektion in Projektion, Selektion in Projektion, Projektion in Selektion, Selektion in Selektion) ist erlaubt.

Algebren für spezielle geschachtelte Relationen

Wir hatten als relevanten Spezialfall in diesem Abschnitt die PNF-Relationen diskutiert. PNF-Relationen haben in jeder (Teil-)Relation einen flachen Schlüssel (also atomare Attribute als Schlüssel). Die Beispielrelation aus Abbildung B.42 ist eine PNF-Relation. Eine Algebra für PNF-Relationen muss bei allen Operationen diese PNF-Eigenschaft erhalten.

Insbesondere haben wir die folgenden Forderungen:

- Die Projektion muss die flachen Schlüssel bewahren. Alternativ muss ein neuer flacher Schlüssel aus den verbleibenden atomaren Attributen gebildet werden; in diesem Fall müssen Tupel verschmolzen werden, wie wir es bei der Vereinigung ausführlich diskutieren werden. In jedem Fall muss auf jeder Ebene mindestens ein atomares Attribut übrig bleiben.
- Verbund und Vereinigung werden rekursiv, da die Schlüsseleigenschaft Verschmelzen von Tupeln erzwingt, die den gleichen Schlüssel, aber unterschiedliche Unterrelationen haben.

Wir zeigen das rekursive Verschmelzen von Tupeln anhand der bereits bekannten Beispielrelation aus Abbildung B.42 von Seite 791.

◀**Beispiel B-6**▶ Abbildung B.46 zeigt eine zweite Relation, die wir mit der bereits vorhandenen vereinigen wollen.

Region	Weingüter			
	Weingut	Anbaugebiet	Weine	
			WeinID	Name
Bordeaux	Château La Rose	Saint-Émilion	9999	Chardonnay
Bordeaux	Magnifique	Pomerol	5588	Beaujolais

Abbildung B.46: Zweite geschachtelte Relation mit demselben Schema wie die Relation in Abbildung B.42

Die Vereinigung dieser beiden geschachtelten Relationen ergibt mit PNF-Vereinigung die in Abbildung B.47 angegebene Relation. Die Tupel der zweiten Relation werden mit dem letzten Tupel der ursprünglichen Relation verschmolzen, indem auf allen Ebenen von oben nach unten eine Vereinigung durchgeführt wird. Dabei wird sowohl ein neuer Wein bei einem existierenden Weingut (Château La Rose) eingefügt, als auch ein neues Weingut mit einem erzeugten Wein aufgenommen. Die beiden neuen Tupel werden also auf unterschiedlichen

Ebenen integriert, da die erste Stufe (Region Bordeaux) jeweils vorhanden ist und die zweite Stufe (das Weingut) vom ersten Tupel ebenfalls vorhanden ist.

Region	Weingüter			
	Weingut	Anbaugebiet	Weine	
			WeinID	Name
Kalifornien	Helena	Napa Valley	3456 3478	Zinfandel Pinot Noir
	Bighorn	Napa Valley	4961	Chardonnay
Bordeaux	Château La Rose	Saint-Émilion	1042 9999	La Rose Grand Cru Chardonnay
	Château La Pointe	Pomerol		
	Magnifique	Pomerol	5588	Beaujolais

Abbildung B.47: Ergebnis der Vereinigung der Relationen in Abbildung B.42 und B.46 als PNF-Relationen

□

Begriff	Informale Bedeutung
NF ² -Relationen	geschachtelte Relationen, nicht mehr in erster Normalform
PNF-Relationen	geschachtelte Relationen in Partitioned Normal Form, die sich vollständig mit Relationen in erster Normalform darstellen lassen
NF ² -Algebra	Erweiterung der Relationenalgebra, die auf geschachtelten Relationen arbeitet

Tabelle B.5: Wichtige Begriffe zu geschachtelt relationalen Modellen und Anfragen

B.3.4 Zusammenfassung

Das Relationenmodell kann um geschachtelte Relationen erweitert werden. Schachtelung bedeutet dabei, dass statt eines elementaren Attributwertes wieder eine gesamte Tabelle als Attributwert eingeführt werden kann. Ein Attribut ist dann selbst wieder ein eigenes Relationenschema.

Zusätzlich sind dann wiederum in passenden Algebren Operationen hinzuzufügen, die die Schachtelung behandeln können. Wir haben verschiedene Erweiterungen der Relationenalgebra in diesem Abschnitt eingeführt.

B.3.5 Vertiefende Literatur

Eine orthogonale Algebra für geschachtelte Relationen wird in [SS86] eingeführt. Lehrbücher zu objektorientierten Datenbankmodellen, -sprachen und -systemen wie [Heu97, SST97, LV95] enthalten auch Abschnitte zu geschachtelten Relationen.

B.3.6 Übungsaufgaben

Übung B-1 Diskutieren Sie eine Modellierung des im Anhang A.2 aufgeführten relationalen Schemas in geschachtelten Relationen.

Welcher Bezug besteht zwischen Fremdschlüsselbedingungen in flachen Relationen und möglichen Modellierungen in geschachtelten Relationen? □

Übung B-2 Im zugrundeliegenden Lehrbuch wurde das laufende Wein-Beispiel um Rebsorten mit ihren Prozentanteilen ergänzt. Diese können auch geschachtelt bei Weinen gespeichert werden. Geben Sie eine geschachtelte Relation an, die dies leistet. Geben Sie zusätzlich folgende Anfragen in einer der obigen Algebren an:

- Alle Weine, die Riesling und Silvaner enthalten.
- Welche zwei Weine eines Erzeugers haben dieselben Prozentanteile? □