 Közlemények 31/1984, 95-113

ÜBER N-FACHE DEKOMPOSITION EINER RELATION IM CODDSCHEN RELATIONENMODELL

LE TIEN VUONG
Institut für Kybernetik und
Rechentechnik, Hanoi, Vietnam

Abstract

This paper introduces a concept: n-ary decomposition of a relation over a set of attributes. The inference rules -System of n-ary Decompositionstructure of this relation and some important properties of a full family of all decompositions have been shown. The concept of antiroot is introduced as a tool for describing the families of decompositions. The necessary and sufficient condition for a relation to be decomposable into n projections is provided in general case.

1. Einleitung

Das Konzept der Dekomposition einer Relation ist sehr wichtig im Prozess des Datenbankentwurfs. Eine Relation $R(U)$ auf einer Menge der Attributen $U=\{A_1,\dots,A_n\}$ wird oft in kleinere Relationen $R_1(X_1),\dots,R_n(X_n)$ mit $X_i \subseteq U$, $\bigcup_{i=1}^{n} X_i=U$ dekomponiert. Bei [COOD 70], [CODD 71] wird dieser Dekompositionsprozess durch die Normalisierung realisiert. Sie basiert auf den Abhängigkeiten zwischen Attributen einer Relation. Die Dekomposition wird informationsverlustsfrei genannt, wenn aus ihren Projektionen durch die natürlichen Verbundoperation die Relation R auf U wiederaufgebaut werden kann.

In der Praxis ist die Bedingung für die informationsverlustfreie Dekomposition der Relation sehr streng. Auf Grund der funktionalen und mehrwertigen Abhängigkeit haben [FAGI 77], [ARDE 79] eine notwendige und hinreichende Bedingung für die binaere Dekomposition einer Relation in 2 Projektionen bewiesen.

In [NICO78], [MEMA 79], [LE 83] haben die Autoren für die Dekomposition einer Relation in 3 Projektionen ohne Informationsverlust auf Grund der gegenseitigen Abhängigkeit gezeigt. Mit Hilfe des verlustfreien Verbundes haben [ABU 79], [BEVA 79], [RISS 77] die Dekomposition einer Relation in mehr als 3 Projektionen betrachtet.

In dieser Arbeit werden die Resultate von [ARDE 79], [LE 83] verallgemeinert. Eine Relation R auf U wird in n (>3) Projektionen ohne Informationsverlust zerlegt. Diese Dekomposition wirn n-fache Dekomposition (Abk. n-Dekompostion) der Relation genannt. Die notwendige und hinreichende Bedingung für die Existenz der n-Dekomposition einer Relation R auf U sowie die Eigenschaften dieser Dekomposition werden gezeigt.

2. Grundbegriffe

Es sei $U = \{A_1,\ldots,A_n\}$ eine endliche Menge, deren Elemente A_i, $i=\overline{1,n}$ Attribute sind. Jedes Attribut A_i hat einen entsprechenden Werte-bereich (Domain) $dom(A_i)$. Im folgenden werden die Buchstaben $A,B,\ldots$ fuer die einzelnen Attribute und $X,Y,\ldots$ für die Menge von Attributen benutzt. Eine Relation R auf U ist eine Untermenge des Kartesischen Produktes der zugehöriger Wertebereiche bzw. Wertemengen

$$R \subseteq dom(A_1)x\ldots x\ dom(A_n).$$

Es wird also eine Relation mit $R(A_1,\ldots,A_n)$ oder $R(U)$ bezeichnet. Es sei t ein Tupel von R. Dann wird mit $t[A_i]$ der Wert von t für das Attribut A_i bezeichnet. Die Projektion von t in $X\subseteq U$ wird mit $t[X]$ geschrieben. Sie ist eine Abbildung aller Attribute aus X in ihre Wertebereiche und wird X-Werte on t genannt.

Die funktionale Abhängigkeit (Abk. FA) $X\rightarrow Y$ ist in einer Relation R erfüllt, wenn aus $t_1[X]=t_2[X]$ für 2 beliebige Tupel $t_1,t_2\in R$ auch $t_1[Y]=t_2[Y]$ folgt (vgl./ARMS 74/). Eine mehrwertige Abhängigkeit (Abk. MWA) $X \to\to Y$ ist in einer Relation $R(U)$ erfüllt, wenn für jedes Paar von Tupeln $t_1, t_2 \in R$ mit $t_1[X]=t_2[X]$ immer ein Tupel $t \in R$ existiert, so dass $t[X \cup Y]=t_1[X \cup Y]$ und $t[U \setminus (X \cup Y)]=t_2[U \setminus (X \cup Y)]$ gelten. Die gegenseitige Abhängigkeit (Abk. GA) $g(X,Y,Z)$ gilt in einer Relation R mit $X \subseteq U$, $Y \subseteq U$, $Z \subseteq U$ und $X \cup Y \cup Z=U$, wenn für je 3 beliebige Tupel $t_1, t_2, t_3 \in R$ mit $t_1[X]=t_2[X]$, $t_2[Y]=t_3[Y]$, $t_3[Z]=t_1[Z]$ immer ein Tupel $t \in R$ existiert, so dass $t[X]=t_1[X]$, $t[Y]=t_2[Y]$, $t[Z]=t_3[Z]$ gelten (vgl. [NICO 78], [MEMA 79]).

In der vorgelegten Arbeit werden nur 2 relationale Operationen, und zwar Projektion und natürlicher Verbund benutzt. Es wird im folgenden die Vereinigung von 2 Mengen $X \cup Y$ in Form XY vereinbart.

3. n-fache Dekomposition

In diesem Abschnitt wird die informationsverlustfreie Dekomposition einer Relation auf der Menge der Attribute in n Projektionen $(n>3)$ betrachtet. Der Begriff der binaeren Dekomposition und der ternaeren Dekomposition (vgl. [ARDE 79], 83]) [LE 83]) wird wie folgt verallgemeinert.

Defintion 3.1

Es sei R eine Relation auf U. Eine n-fache Dekomposition von R ist ein n-Tupel $(X_1,\dots,X_n)$ der Untermengen von U mit $X_i \subseteq U$, $\bigcup_{i=1}^{n} = U$, so dass für beliebige Tupel $t_i \in R$, $i=\overline{1,n}$

$$t_i[X_i \cap X_j] = t_j[X_i \cap X_j], \quad i \neq j, \quad i,j=\overline{1,n}$$

gelten, und ein Tupel $t \in R$ existiert, wobei $t[X_i]=t_i[X_i]$ $i=\overline{1,n}$ ist.

Es werden die Menge $W= \bigcup_{i \neq j} S_{ij}$ mit $S_{ij}=X_i \cap X_j$ die Wurzel der n-Dekomposition, $Z_i=X_i \setminus W$ X_i-Zweig und $W_i=X_i \cap W$ X_i-Wurzel

$i=\overline{1,n}$ der Dekomposition genannt. $X_i, i=\overline{1,n}$ sind Komponenten. Alle möglichen n-Dekompositionen einer Relation werden in eine Familie der Dekompositionen zusammengefasst. Diese Familie wird Dekompositionsstruktur $\mathcal{D}$ genannt. Dann werden die Eigenschaften der Dekompositionsstruktur der Relation R auf U im folgenden Satz formuliert.

Satz 3.1

Es sei $\mathcal{D}$ eine Familie der n-Dekompositionen einer Relation R auf U. Dann erfüllt $\mathcal{D}$ die folgenden Bedingungen:

D1. $(\emptyset,\ldots,\emptyset,U)\in\mathcal{D}$

D2. Wenn $(X_1,\ldots,X_n)\in\mathcal{D}$ ist, sind auch $(X_{\pi(1)}, X_{\pi(2)},\ldots,X_{\pi(n)})\in\mathcal{D}$, wobei $\pi:\{1,2,\ldots,n\}\to\{1,2,\ldots,n\}$ eine Permutation ist.

D3. Wenn $(X_1,\ldots,X_n)\in\mathcal{D}$, $X_i \subseteq Y \subseteq U$, $i=\overline{1,n}$ ist, dann gilt $(X_1,\ldots,X_{i-1},Y,X_{i+1},\ldots,X_n)\in\mathcal{D}$.

D4. Wenn $(X_1,\ldots,X_n)\in\mathcal{D}, X_i \subseteq X_j$, $i\neq j$ sind, ist auch $(X_1,\ldots,X_{i-1},\emptyset,X_{i+1},\ldots,X_j,\ldots,X_n)\in\mathcal{D}$.

D5. Wenn $(X_1,\ldots,X_n),(Y_1,\ldots,Y_n)\in\mathcal{D}$ mit $Y_1\cap Y_i$, $i=\overline{2,n}$ $Y_i\cap Y_j \subseteq X_i\cap X_j$, $i\neq j$, $i,j\geq 2$ sind, dann gilt $(X_1\cap Y_1, Y_2,\ldots,Y_n)\in\mathcal{D}$.

Beweis

Die Bedingungen D1 bis D4 folgen ohne Schwierigkeit aus der Definition 3.1. Hier wird nur D5 bewiesen.

Für D5 werden die Bedingungen aus der Definition 3.1 geprüft.

a. Da $Y_1\cap Y_i=X_i$, $i=\overline{2,n}$ sind, gilt dann

$$(X_1\cap Y_1)\cup\bigcup_{i=2}^{n} Y_i=(X_1\cup\bigcup_{i=2}^{n} Y_i)\cap(\bigcup_{i=1}^{n} Y_i)\supseteq(\bigcup_{i=1}^{n} X_i)\cap(\bigcup_{i=1}^{n} Y_i)=U.$$

b. Da $(X_1,\ldots,X_n)\in\mathcal{D}$ ist, existieren $t_i\in R, i=\overline{1,n}$ und $t'\in R$ mit $t_i[X_i\cap X_j]=t_j[X_i\cap X_j]$, $i\neq j$, $i,j=\overline{1,n}$ und $t'[X_i] = t_i[X_i]$, $i=\overline{1,n}$. Da $Y_1\cap Y_i=X_i$, $Y_i\cap Y_j \subseteq X_i\cap X_j$, $i\neq j$, $i,j=\overline{2,n}$

sind, sind für $(Y_1,\ldots,Y_n)\in\mathcal{D}$ zugleich $t', t_i\in R$, $i=\overline{2,n}$ und $t\in R$, die die Definition 3.1 erfüllen. Tatsächlich gelten $t'[Y_1\cap Y_j]=t_j[Y_1\cap Y_j]$, $j=\overline{2,n}$ und $t_i[Y_i\cap Y_j]=t_j[Y_i\cap Y_j]$, $i\neq j$, $i,j=\overline{2,n}$, und $t[Y_1]=t'[Y_1]$, $t[Y_i]=t_i[Y_i]$, $i=\overline{2,n}$. Zusammen mit $t_1[X_1]=t'[X_1]$ $t[Y_1]=t'[Y_1]$ gilt, das auf $X_1\cap Y_1$ $t=t'=t_1$ ist und alle Bedingungen der Definition 3.1 erf werden. So ist

$$(Y_1\cap Y_1,Y_2,\ldots,Y_n)\in\mathcal{D}\ .$$

Folgerung 3.1

D6. Ist eine der $X_i, i=\overline{1,n}$ gleich U, dann $(X_1,\ldots,X_n)\in\mathcal{D}$.

D7. Wenn $(X_1,\ldots,X_n)\in$ ist, gilt dann auch
$(X_1,\ldots,X_{i-1},\emptyset,X_{i+1},\ldots,X_i\cup X_j,\ldots,X_n)\in\mathcal{D}$.

Beweis

D6 werden D1 und D3 und für D7 die Bedingungen D2, D3 und D4 zum Beweis benutzt. Daraus ergibt sich die Behauptung.

Um das im Satz 3.1 angegebene Eigenschaften-System als vollständiges System zu zeigen, wird eine vollständige Familie der n-Dekompositionen (analog wie vollständige Familie der funktionalen Abhängigkeiten (vgl. [ARMS 74])) wie folgt definiert:

Definition 3.2

Es sei U eine Menge der Attribute. Eine Familie $\mathcal{D}$ von allen n-Tupeln $(X_1,\ldots,X_n)$ der Untermengen von U mit $\bigcup_i X_i=U$, wobei sie alle Bedingungen D1 bis D5 erfüllen, wird eine vollständige Familie der n-Dekompositionen auf U genannt.

Nach der Definition 3.2 kann der Satz 3.1 wie folgt umformuliert werden: Es sei $\mathcal{D}$ eine Familie aller n-Dekompositionen der Relation R auf U. Dann ist $\mathcal{D}$ eine vollständige Familie der n-Dekompositionen auf U.

Damit einige wichtige Eigenschaften der vollständigen Familie der n-Dekompositionen untersucht werden können, wird der "duale" Begriff, nämlich die "Nichtdekomposition", Eingeführt. Dann gilt der folgende Satz.

Satz 3.2

Es sei $\mathcal{D}$ eine vollständige Familie der n-Dekompositionen der Relation R auf U. Dann gelten folgende Eigenschaften:

ND1. Wenn $(X_1,\ldots,X_n)\notin\mathcal{D}$, $\bigcup_{i=1}^{n} X_i=U$ sind, sind alle $X_i, i=\overline{1,n}$ keine ineinander geschachtelten Untermengen von U.

ND2. Wenn $\bigcup_{i=1}^{n} X_i=U$, $i=\overline{1,n}$, $(X_1,\ldots,X_nY)\notin\mathcal{D}$, $Y\subseteq U$ sind, so ist $(X_1,\ldots,X_n)\notin\mathcal{D}$.

ND3. Wenn $\bigcup_{i=1}^{n} X_i=U$, $i=\overline{1,n}$, $(X_1,\ldots,X_n)\notin\mathcal{D}$ ist, existiert eine Untermenge $S\subseteq U$, wobei $X_i\cap X_j\subseteq S$, $i\neq j$, $i,j=\overline{1,n}$, und mindestens 2 Untermengen von $\{X_1,\ldots,X_n\}$, nämlich X_p,X_q mit $X_p\nsubseteq S$, $X_q\nsubseteq S$ sind. Umgekehrt gilt fuer jedes $(Y_1,\ldots,Y_n)$ mit $\bigcup_{i=1}^{n} Y_i=U$, $Y_i\cap Y_j\subseteq S$, $i\neq j$, $i,j=\overline{1,n}$ und mindestens zwei Untermengen $Y_p,Y_q\subseteq U$ mit $Y_p\nsubseteq S$, $Y_q\nsubseteq S$ immer $(Y_1,\ldots,Y_n)\notin\mathcal{D}$.

Beweis

ND1 ist ohne Schwierigkeit einzusehen.

Zu ND2:

Es seien $\bigcup_{i=1}^{n} X_i=U$ und $\forall X_i\neq\emptyset$. Dann gilt entweder $(X_1,\ldots,X_n)\in\mathcal{D}$ oder $(X_1,\ldots,X_n)\notin\mathcal{D}$. Angenommen ist es $(X_1,X_2,\ldots,X_n)\in\mathcal{D}$. Aus D3 folgt $(X_1,\ldots,X_nY)\in\mathcal{D}$, wobei $Y\subseteq U$ ist. Dies ist ein Widerspruch zur Voraussetzung. Deshalb muss $(X_1,\ldots,X_n)\notin\mathcal{D}$ sein.

Zu ND3:

Es wird die Familie $\mathcal{B}$ von den Mengen $S\subseteq U$ betrachtet, so dass $X_i\cap X_j\subseteq S$, $i\neq j$, $i,j=\overline{1,n}$, und $(X_1S,\ldots,X_nS)\notin\mathcal{D}$ sind.

Im Spezialfall ist auch $X_i \cap X_j = S$ in der Familie. Wegen der Endlichkeit der Menge U kann eine Menge $S' \in \mathcal{B}$ gewaehlt werden, die keine Untermenge von irgendeiner anderen in $\mathcal{B}$ ist., d.h. S ist eine maximale Untermenge. Die X_i-Wurzel, $\overline{i=1,n}$ können wie möglich weiter vergrössert werden, bis die Resultate noch immer nicht zu $\mathcal{D}$ gehoeren zu bleiben. Jetzt muss gezeigt werden, dass S' die Bedingung ND3 erfüllt.
O.B.d.A. wird vorausgesetzt, dass $X_i \subseteq S'$, $i=1,\overline{n-2}$, d.h. $n-2$ erste Untermengen X_i in S' enthalten sind. Weiter sei $X_{n-1} \subseteq S'$ Dann gilt $X_n S'=U$. So gilt nach D6 $(X_1,\ldots,X_n S') \in \mathcal{D}$. Das ist ein Widerspruch. Das bedeutet, dass nicht gleichzeitig $n-1$ Untermengen von $\{x_1,\ldots,x_n\}$ in S' enthalten sind. Daraus folgt $X_{n-1} \nsubseteq S'$. Ganz analog wird auch für $X_n \nsubseteq S'$ bewiesen.
Umgekehrt seien $(Y_1,\ldots,Y_n)$ mit $Y_i \cap Y_j \subseteq S'$, $i \neq j$, $\bigcup_{i=1}^{n} Y_i=U$. Es existieren mindestens zwei Untermengen von $\{Y_1,\ldots,Y_n\}$, nämlich $Y_k \nsubseteq S'$ und $Y_m \nsubseteq S'$. Dann muss gezeigt werden, dass $(Y_1,\ldots,Y_n) \notin \mathcal{D}$ ist. Angenommen $(Y_1,\ldots,Y_n)$, die die obigen Bedingungen erfuellt, aber $(Y_1,\ldots,Y_n) \in \mathcal{D}$. Es muss gezeigt werden, dass die Maximalität von S' verletzt wird. Zuerst wird dies für den Fall. $Y_i \cap Y_j = S'$, $i \neq j$ bewiesen, dann folgt mit Hilfe von ND2 der allgemeine Fall.

Aus $(Y_1,\ldots,Y_n) \in \mathcal{D}$ folgt mit mehreren Anwendungen von D3, D2:

$$(X_1Y_1,\ldots,X_nY_n) \in \mathcal{D} ,$$
$$(X_1Y_2,\ldots,X_nY_{n-1}) \in \mathcal{D} ,$$
$$\ldots\ldots\ldots$$
$$(X_1Y_n,\ldots,X_nY_1) \in \mathcal{D} .$$

Aus den letzten n Ausdrücken und aus $(Y_1,\ldots,Y_n) \in$ folgen mit mehreren Anwendungen von D5 und D2

$$(X_1Y_1 \cap X_2,\ldots,X_nY_n \cap Y_{n-1}) \in \mathcal{D} ,$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$(X_1Y_n \cap Y_{n-1},\ldots,X_nY_1 \cap Y_2) \in \mathcal{D} .$$

Aus diesen n Ausdruecken und mit Anwendungen von D3 und den Bedingungen $Y_i \cap Y_j = S'$, $i \neq j$ gilt nach den ausfuehrlichen Umrechnungen aller Komponenten

$$(X_1 S', \ldots, X_n S') \in \mathcal{D} .$$

Das ist ein Widerspruch zur Definition der Menge S'. So muss

$$(Y_1, \ldots, Y_n) \in \mathcal{D}$$

sein. Zusammen mit ND2 gilt dann für auch allgemeine Fall.

Es ist notwendig, eine notwendige und hinreichende Bedingung fuer die Existenz einer n-Dekomposition einer Relation zu finden. Zu diesem Zweck wird der Begriff der Antiwurzel der vollständigen Familie der gen Familie der Dekomposition einge-Dekomposition eingeführt (vgl. [ARDE 79],[LE 83]). Die Antiwurzel wird wie folgt definiert:

Definition 3.3

Es sei $\mathcal{D}$ eine vollständige Familie der n-Dekompositionen auf U. Die nichttriviale Antiwurzel von $\mathcal{D}$ ist die Menge $S \subseteq U$ mit $card(U \setminus S) \geq 2$, wobei für jedes $(X_1, \ldots, X_n)$, $X_i \subseteq U$, $i = \overline{1, n}$, $\bigcup_{i=1}^{n} X_i = U$, die Bedingungen $X_i \cap X_j \subseteq S$, $i \neq j$, für mindestens zwei Mengen, z.B. $X_p \not\subseteq S$, $X_q \not\subseteq S$, $p, q \in \{1, 2, \ldots n\}$ gelten, dann ist $(X_1, \ldots, X_n) \notin \mathcal{D}$.

Alle Untermengen $S \subseteq U$, die die obigen Bedingungen mit $card(U \setminus S) < 2$ erfüllen, werden triviale Antiwurzeln genannt.

$\mathcal{A}$ bezeichnet die Familie der nichttrivialen Antiwurzeln von $\mathcal{D}$ auf U. Dann gilt der

Satz 3.3

$\mathcal{D}$ kann vollstaendig mit ihrer Familie $\mathcal{A}$ beschrieben werden. Das bedeutet:

Jedes $(X_1,...,X_n)$ mit $\bigcup_{i=1}^{n} X_i=U$, $X_i \subseteq U$ gehört genau dann $\mathfrak{D}$, wenn für jedes $S\in\mathfrak{A}$ mit $X_i \cap X_j \subseteq S$, $i\neq j$, $i,j=\overline{1,n}$ mindestens $n-1$ dieser Untermengen (von $\{X_1,...,X_n\}$) in S enthalten sind.

Beweis

Die Behauptung gilt analog zu dem folgenden: $(X_1,...,X_n) \notin\mathfrak{D}$ ist genau dann erfüllt, wenn eine $S\in\mathfrak{A}$ existiert, wobei $X_i\cap X_j \subseteq S$, $i\neq j$, $i,j=\overline{1,n}$, von denen mindestens zwei nicht in S enthalten sind.

Dies wird mit ND3 und der Definition 3.3 beweisen.

4. Beziehungen zwischen Abhängigkeitsstruktur und n-Dekomposition

Die Beziehungen zwischen den Datenabhängigkeitsarten und binären bzw. ternären Dekompositionen wurden in [RISS 77], [ARDE 79], [LE 83] diskutiert.

Es ist wohlbekannt, dass die Menge $\mathfrak{A}$ der nichttrivialen Antiwurzeln der Dekompositionsstructur von R auf U eine Untermenge der abgeschlossenen Menge $\mathfrak{B}$ der Attribute bzgl. der funktionalen Abhängigkeitsstructur von R ist. (Im Armstrong-Sinne, vgl. [ARMS 74]), d.h. $\mathfrak{A} \subseteq \mathfrak{B}$ (vgl. [ARDE 79], [LE 83]).
Die Eigenschaften der Menge $\mathfrak{B}$ bzw. die Anzahl der maximalen Elemente der zugehoerigen FD-Structur F können ausführlich in [ARMS 74], [BÉKÉ 80] gefunden werden. Es ist nicht schwierig zu zeigen, dass alle nichttriviale Antiwurzeln der n-Dekompositionsstructur $\mathfrak{D}$ der Relation R auf U eine abgeschlossene Menge bilden.

Mit Hilfe der Resultate im obigen Abschnitt kann jede n-Dekompositionsstructur $\mathfrak{D}$ der Relation R durch ihre entsprechende Menge der Antiwurzeln $\mathfrak{A}$ beschrieben und aufgebaut werden.

Das Ziel dieses Abschnittes besteht darin, einige Eigenschaften der Antiwurzeln der n-Dekompositionen zu studieren und den Spezialfall der n-Dekomposition mit leeren Wurzeln zu entwickeln.

Es gilt der folgende

Hilfssatz 4.1

Es sei $(X_1,\ldots,X_n)\in\mathcal{D}$ eine n-Dekomposition der Relation R auf U. Dann sind alle Komponenten X_i, $i=\overline{1,n}$ die abgeschlossenen Mengen bzgl. FA-Structur dieser Relation.

Beweis

Es sei $(X_1,\ldots,X_n)\in\mathcal{D}$. Nach dem Satz 3.3 existiert eine Menge $S\subseteq U$ mit $X_i\cap X_j\subseteq S$, $i\neq j$, $i,j=\overline{1,n}$ und mindestens $n-1$ Untermengen X_i mit $X_i\subseteq S$. O.B.d.A. wird $X_i\subseteq S$, $i=\overline{1,n-1}$ vorausgesetzt. Für den Beweis ist es ausreichend $X_i\cap X_j=S$ und ND2 zu benützen. Zuerst wird für X_1 beweisen, dann gilt es auch für all X_i, $i=\overline{1,n}$.

Es muss gezeigt werden, dass $X_1\not\to A$ ist, wobei A ein Attribut aus U, $A\notin X_1$ ist. Dann muss $A\in X_j\setminus X_1$ bei festem $j\in\{2,\ldots,n\}$ sein.

Es seien zwei beliebige Tupel $t_1, t'\in R$ ausgewählt, so dass $t_1[X_1]=t'[X_1]$ aber $t_1[A]\neq t'[A]$ sind. Da S abgeschlossen und $X_i\cap X_j=S$ ist, können zwei Tupel $t_1, t_j\in R$ (für bestimmtes j) ausgewählt werden, wobei $t_1[X_1\cap X_j]=t_j[X_1\cap X_j]$ und $t_1[A]\neq t_j[A]$, $A\in X_j\setminus X_1$ sind.

Wegen $(X_1,\ldots,X_n)\in\mathcal{D}$ gilt für beliebige $t_i\in R$, $i=\overline{1,n}$ und $t'\in R$: $t_i[X_i\cap X_j]=t_j[X_i\cap X_j]$ und $t'[X_i]=t_i[X_i]$. Bei festem j gelten auch $t'[X_j]=t_j[X_j]$ und $t'[A]=t_j[A]$. Dann gilt

$$t'[X_1]=t_1[X_1] \textit{ aber } t_1[A]\neq t'[A]=t_j[A].$$

Das bedeutet, dass X_1 abgeschlossen ist. Ganz analog gelten für

alle X_i, $i=\overline{1,n}$.
Zusammen mit ND2 gilt die Behauptung des Hilfssatzes.

Satz 4.2

Es seien R eine Relation auf U, S eine abgeschlossene Menge bzgl. der FA-Structur von R, so dass S nicht durch den Durchschnitt von zwei von S verschiedenen abgeschlossenen Mengen generiert wird. Es sei $card(U\setminus S)\geq 2$. Dann ist S eine Antiwurzel der n-Dekompositionsstructur von R.

Beweis

Die Behauptung wird mit Hilfe des Hilfssatzes 4.1 bewiesen.

Es ist bekannt, dass die Familie der abgeschlossenen Mengen bzgl. der FA-Struktur für die Durschschnittsoperation abgeschlossen itst, aber die Familie der Antiwurzeln nicht. Es gilt also nur: Wenn S_1, S_2 mit $S_1 \subseteq U$, $S_2 \subseteq U$ und $S_1 \cup S_2 \neq U$ abgeschlossen sind, dann ist $S_1 \cap S_2$ eine Antiwurzel (vgl. [ARDE 79], [LE 83]).
Für die Beziehung zwischen Funktionalen bzw. mehrwertigen Abhängigkeiten und n-Dekompositionsstructur der Relation R gelten die folgenden Verallgemeinerungen.

Satz 4.3 ([BEVA 79], [LE 83])

Wenn eine Untermenge in einem X_i-Zweig der n-Dekomposition von einer mit X_i-Zweig disjunkten Untermenge funktional abhängig ist, dann ist sie auch von der entsprechenden X_i-Wurzel abhängig.

Beweis

Der Satz kann wie folgt umformuliert werden:
Es seien $(X_1,\ldots,X_n)\in\mathcal{D}$, $S\rightarrow Y$, $S \subseteq U$, $S\cap Z_{X_i}=\emptyset$, $Y \subseteq Z_{X_i}$ dann gilt $W_{X_i}\rightarrow Z_{X_i}$ (vgl. Absch.3.).
Der Beweis läuft ganz analog wie in [LE 83].

<u>Satz 4.4</u> ([MEMA 79])

Es seien R eine Relation auf U, $X_i \subseteq U$, $\bigcup_{i=1}^{n} X_i = U$. $(X_1,\ldots,X_n)$ gehört genau dann zur n-Dekompositionsstructur $\mathcal{D}$ der Relation R auf U, wenn $(X_1 \cap W,\ldots,X_n \cap W)$ zu $\mathcal{D}_w$ gehört und $W_i \twoheadrightarrow X_i \setminus W$ ist, wobei $\mathcal{D}_w$ die n-Dekompositionsstructur der Relation auf der Menge $W = \bigcup_{i \neq j} (X_i \cap X_j)$, $W_i = X_i \cap W$ ist.

<u>Beweis</u>

Angenommen $(X_1,\ldots,X_n) \in$, $\bigcup_{i=1}^{n} X_i = U$, $W = \bigcup_{i \neq j} (X_i\ X_j)$, $i,j=\overline{1,n}$, $W_i = X_i \cap W$. Dann existieren $t_i \in R$, $i=\overline{1,n}$ und $t \in R$, wobei

$$t_i[X_i \cap X_j] = t_j[X_i \cap X_j],$$
$$t[X_j] = t_i[X_i]$$

sind.

Da $W_i \cap W_j = (X_i \cap W) \cap (X_j \cap W) = (X_i \cap X_j) \cap W \subseteq X_i \cap X_i$, und $W_i \subseteq X_i$ sind, gilt zugleich

$$t_i[W_i \cap W_j] = t_j[W_i \cap W_j] \quad und$$
$$t[W_i] = t_i[W_i].$$

Deshalb gilt die n-Dekomposition der Relation auf der Menge $W = \bigcup_{i=1}^{n} W_i = \bigcup_{i=1}^{n} (X_i\ W)$. Mit anderen Worten:

$$(X_1 \cap W,\ldots,X_n \cap W) \in \mathcal{D}_w.$$

Auf Grund der binären Dekomposition [ARBE 79] gilt für $(X_1,\ldots,X_n) \in \mathcal{D}$:

$$(X_1, \bigcup_{i=1}^{n} X_i, \emptyset,\ldots,\emptyset) \in \mathcal{D} \quad und$$

$$(\bigcup_{i=2}^{n} X_i)\ X_1 \twoheadrightarrow X_1$$ (vgl.[FAGI 77], [ARDE 79]).

Wegen $(\bigcup_{i=2}^{n} X_i)\cap X_1=W\cap X_1=W_1$ und $Z_1=X_1\setminus W \subseteq X_1$ gilt dann $W_1\twoheadrightarrow Z_1$ oder $W_1\twoheadrightarrow X_1\setminus W$.

Es gilt analog für alle $i, i=\overline{1,n}$, d.h. $W_i\twoheadrightarrow X_i\setminus W$.
Umgekehrt sei angenommen, dass $(X_1,\ldots,X_n)$ mit $\bigcup_{i=1}^{n} X_i=U$ und $W=\bigcup_{i\neq j}(X_i\cap X_j)$, $W_i=X_i\cap W$, $Z_i=X_i\setminus W_i$, $(W_1,\ldots,W_n)\in\mathcal{D}_w$ und $W_i\twoheadrightarrow Z_i$ gelten, so muss gezeigt werden, dass $(X_1,\ldots,X_n)\in\mathcal{D}$ ist.

Tatsächlich existieren $t_i\in R$, $i=\overline{1,n}$ und $t\in R$ mit

$$t_i[W_i\cap W_j]=t_j[W_i\cap W_j] \quad \text{und}$$
$$t[W_i]=t_i[W_i].$$

Dann folgt wegen

$$W_i\cap W_j=X_i\cap X_j \quad W=X_i\cap X_j$$
$$t_i[X_i\cap X_j]=t_j[X_i\cap X_j]. \quad (+)$$

Aus $W_1\twoheadrightarrow Z_1$ gelten nämlich für $t_1,t_k\in R$:

$$t_k[W_1]=t_1[W_1] \quad \text{oder}$$
$$t_k[W\cap X_1]=t_1[W\cap X_1] \text{ und } t'\in R$$
$$t'[W_1Z_1]=t_1[W_1Z_1] \text{ oder } t'[X_1]=t_1[X_1] \text{ und}$$
$$t'[W_1(U\setminus X_1)]=t_k[W_1(U\setminus X_1)].$$

Es gilt also für alle $i=\overline{1,n}$

$$t'[X_i]=t_i[X_i] \quad (++)$$

Aus (+) und (++) ergibt sich die Behauptung.

Wie oben gezeigt, die Familie der n-Dekompositionen einer Relation wird durch die Familie der entsprechenden Antiwurzeln charakterisiert. Hier wird der Spezialfall mit der leeren Wurzel, d.h. die Menge W=$\emptyset$, betrachtet.

Gegeben ist eine vollständige Familie der n-Dekompositionen auf U. Sie wird in mehrere Teile X_i, $i=\overline{1,n}$ zerlegt. Auf jedem X_i wird eine neue Familie der Dekompositionen $\mathcal{D}_{X_i}$ wieder definiert. Es müssen neue Relationen R_i auf X_i konstruiert werden. Dann gilt der folgende

Satz 4.5

Es sei $\mathcal{D}$ eine vollständige Familie der Dekompositionen auf U und $(X_1,\ldots,X_n)\in\mathcal{D}$, wobei $X_i\cap X_j=\emptyset$, $i\neq j$ $i,j=\overline{1,n}$, $\bigcup_{i=1}^{n} X_i=U$ sind. Dann sind

$$\mathcal{D}_{X_i}=\{(V_1\ X_i,\ldots,V_n\ X_i)/(V_1,\ldots,V_n)\in\mathcal{D}\},\ i=\overline{1,n}$$

die vollständigen Familien der Dekompositionen auf X_i.

Beweis

D1 bis D4 sind ohne Schwierigkeiten zu zeigen.

Zu D5:

Es seien für jedes $i=\overline{1,n}$ $(X_{i1},\ldots,X_{in})\in\mathcal{D}_{X_i}$, $(X'_{i1},\ldots,X'_{in})\in\mathcal{D}_{X_i}$ mit $X'_{i1}\cap X'_{ik}=X_{ik}$, $k=\overline{2,n}$

$X'_{ik}\cap X'_{il}\subseteq X_{ik}\cap X_{il}$, $i,l,=\overline{2,n}$, $i\neq l$, $\bigcup_{k=1}^{n} X_{ik}=\bigcup_{k=1}^{n} X'_{ik}=X_i$.

Es muss gezeigt werden, dass $(X_{i1}\cap X'_{i1},X'_{i2},\ldots,X'_{in})\in\mathcal{D}_{X_i}$ ist.

Es existieren $(V_1,\ldots,V_n),(V'_1,\ldots,V'_n)\in\mathcal{D}$ mit

$$X_{i1}=V_1\cap X_i,\ldots,X_{in}=V_n\cap X_i\ ,$$
$$X'_{i1}=V'_1\cap X_i,\ldots,X'_{in}=V'_n\cap X_i\ .$$

Nach D3 kann $\overline{X}_i=U\setminus X_i$ zur Wurzel der Dekomposition in vergrössert werden. Das ist ganz ähnlich mit den Voraussetzungen wie

$$V_1=X_{i1}\overline{X}_i,\ V_2=X_{i2}\overline{X}_i,\ldots,V_n=X_{in}\overline{X}_i,$$

$$V'_1 = X'_{i1}\bar{X}_i,\ V'_2 = X'_{i2}\bar{X}_i, \ldots, V'_n = X'_{in}\bar{X}_i.$$

Dann gelten

$$V'_1 \cap V'_j = X'_{i1}\bar{X}_i \cap X'_{ij}\bar{X}_i = \bar{X}_i(X'_{i1}\cap X'_{ij}) = \bar{X}_i X_{ij} = V_j,\ j=\overline{2,n}$$

$$V'_k \cap V'_l = X'_{ik}\bar{X}_i \cap X'_{il}\bar{X}_i = \bar{X}_i(X'_{ik}\cap X'_{il}) \subseteq V_k\ V_l\ ,\ k,l=\overline{2,n}.$$

Das bedeutet, dass $(V_1\ V'_1, V'_2, \ldots, V'_n) \in \mathcal{D}$ gilt, dann gilt nach der Definition von $\mathcal{D}_{X_i}$ auch

$$((V_1 \cap V'_1)\cap X_i,\ V'_2\cap X_i, \ldots, V'_n \cap X_i) \in \mathcal{D}_{X_i}$$

oder

$$(X_{i1}\cap X'_{i1}, X'_{i2}, \ldots, X'_{in}) \in \mathcal{D}_{X_i}\ ,\ \text{fuer alle}\ \ i=\overline{1,n}\ .$$

Daraus folgt die Behauptung des Satzes.

Satz 4.6

Es seien n vollständige Familien der Dekompositionen $\mathcal{D}_{X_i}$ auf X_i, $i=\overline{1,n}$, wobei $\bigcup_{i=1}^{n} X_i = U$, $X_i\ X_j = \emptyset$, $i \neq j$, $i,j=\overline{1,n}$ sind. Dann ist

$$\mathcal{D} = \{(\bigcup_i X_{i1}, \ldots, \bigcup_i X_{in}) / (X_{i1}, \ldots, X_{in}) \in \mathcal{D}_{X_i}\ ,\ i=\overline{1,n}\}$$

eine vollständige Familien der Dekomposition auf U, die aus $\mathcal{D}_{X_i}$, $i=\overline{1,n}$ mit $(X_1, \ldots, X_n) \in \mathcal{D}$, $X_j = \bigcup_i X_{ij}$, $j=\overline{1,n}$, $X_i \cap X_j = \emptyset$, $i \neq j$, $i,j=\overline{1,n}$, $\bigcup_i X_i = U$ besteht.

Beweis

D1 bis D4 ist nicht schwierig zu prüfen.

Zu D5:

Angenommen $(V_1,\dots,V_n),(V'_1,\dots,V'_n)\in\mathcal{D}$ mit $V'_1\cap V'_i=V_i$, $V'_i\ V'_j\subseteq V_i\cap V_j$, $i\neq j$, $i,j=\overline{2,n}$. Dann sind fuer X_i, $i=\overline{1,n}$

$$(V_1\cap X_i,\dots,V_n\cap X_i)\in\mathcal{D}_{X_i}\ ,$$

$$(V'_1\cap X_i,\dots,V'_n\cap X_i)\in\mathcal{D}_{X_i}$$

Seien $V_k\cap X_i=X_{ik}$, $V'_k\ X_i=X'_{ik}$, für festes i, $i\in\{1,\dots,n\}$, $k=\overline{1,n}$. Dann gelten

$$(X_{i1},\dots,X_{in})\in\mathcal{D}_{X_i},$$

$$(X'_{i1},\dots,X'_{in})\in\mathcal{D}_{X_i}.$$

Es gelten also weiter

$$X'_{i1}\cap X'_{ik} = X_i\cap(V'_1\cap V'_k) = X_i\cap V_k = X_{ik},\ k=\overline{2,n}$$

$$X'_{ik}\cap X'_{il} = X_i\cap(V'_k\cap V'_l) \subseteq X_{ik}\cap X_{il},\ k\neq l,\ k,l=\overline{2,n}\ .$$

Die obigen Ausdruecken bedeuten, dass alle Bedingungen aus D5 für $\mathcal{D}_{X_i}$ erfüllt sind. So sind

$$(X_{i1}\cap X'_{i1},X'_{i2},\dots,X'_{in})=((V_1\cap X_i)\cap(V'_1\cap X_i),\dots,V'_n\cap X_i)\in\mathcal{D}_{X_i}.$$

Nach der Definition von $\mathcal{D}$ gilt

$$((\bigcup_i X_{i1})\cap(\bigcup_i X'_{i1}),\ \bigcup_i X_{i2},\dots,\bigcup_i X_{in}) =$$

$$(\bigcup_i[(V_1\cap X_i)\cap(V'_1\cap X_i)],\dots,\bigcup_i(V'_n\cap X_i)) =$$

$$((V_1\cap V'_1)\cap(\bigcup_i X_i\ ,\dots,V'_n\cap(\bigcup_i X_i))=$$

$$(V_1\cap V'_1,V'_2,\dots,V'_n)\in\mathcal{D}\qquad \textit{wegen}\quad \bigcup_i X_i=U.$$

Ausserdem gelten noch

$$(\emptyset,\dots,X_i,\emptyset,\dots,\emptyset)\in\mathcal{D}_{X_i}\ ,\quad i=\overline{1,n}\ ,\qquad \text{dann gilt}$$

$$(X_1,\ldots,X_n)=(X_1 \cup \emptyset \cup .. \cup \emptyset, \emptyset \cup X_2 \cup \emptyset \cup .. \cup \emptyset, \ldots, \emptyset \cup \ldots \cup X_n) \in \mathcal{D}$$

So gilt die Behauptung.

5. Schlussfolgerung

In dieser Arbeit wurde der Begriff n-fachen Dekomposition einer Relation auf der Menge der Attributen eingefuehrt. Dieser Begriff ist äquivalent zu der n-Join Abhängigkeit. Die vollständige Familie aller möglichen n-Dekompositionen wurde mit Hilfe der entsprechenden Antiwurzeln betrachtet, die im Armstrong-Sinne abgeschlossen sind. Die notwendige und hinreichende Bedingung für ihre Existenz wurden im verallgemeinerten Fall gezeigt. Ein Spezialfall gezeigt. Ein Spezialfall der Dekomposition mit leeren Wurzel ist auch betractet. Mit diesen Resultaten ist es möglich, den Prozess des Datenbankentwurfes ver- zu verbessern.

Der Autor bedankt sich recht herzlich bei .
Prof. J. Demetrovics und Dr. A. Békéssy für die wertvollen Hinweise und Hilfe zu dieser Arbeit bzw. für das Lesen des Manuskriptes.

Literatur

ABU 79 : A.V.Aho,C.Beeri, J.D.Ullman: The Theory of Join in Relational Databases, ACM TODS 4,3,Sept 1979,279-314

ARMS 74 : W.W.Armstrong: Dependency structure of Data Base Relationsihps, Inf. Pro. 74, 580-583.

ARDE 79 : W.W.Armstrong, C.Delobel: Decompositions and functional Dependenceis in Relation, Université de Montreal 1979 Pub. 271

BEVA 79 : C.Beeri, M.Y.Vardi: On the Properties of total Join Dependencies, Nov. 1979.

BÉKÉ 80 : A.Békéssy, J.Demetrovics, L.Hannák, P.Frankl, Gy. Katona: On the number of maximal dependencies in a Data Basa Relation of fixed order, Disc., Math. 30, 1980, 83-88

CODD 70 : E.F.Codd: A relational Model of Data for large schared Data Bank, C.ACM 13, 6, (1970) 377-387

CODD 71 : E.F.Codd: Further normalization of the Data Base relational model, IBM-Res. RJ 909, AUG. 1971

FAGI 77 : R.Fagin: Multivalued dependencies and a new Normal Form for relational Databases, ACM TODS 2,3 (1977), 262-278

LE 83 : Le Tien Vuong: Untersuchung zur ternaeren Dekomposition einer Relation und zur anwndung der unscharfen Mengen im CRM, Diss. TU-Dresden, 1983

MEMA 79 : A.O. Mendelzon, D.Maier: Generalized mutual dependencies and the deomposition of Data Base Relations, Pro. 5th. Conf. VLDB, Rio de Janeiro, oct. 1979, 75-82.

NICO 78 : J.M.Nicolas: Mutual Dependencies and some results on undencomposable Relations, Proc. 4th. Conf. VLDB, Berlin, sept. 1978, 360-367

RISS 77 : J.Rissaenen: Independent components of Relations ACM TODS 2,4 (1977) 317-325 .

Ö S S Z E F O G L A L Ó

A CODD-FÉLE RELÁCIÓS-MODELLBEN LÉVŐ RELÁCIÓ N-SZERES DEKOMPOZICIÓJÁRÓL

Le Tien Vuong

A cikkben a szerző levezeti az n-szeres dekompozició fogalmát és a dekompoziciók teljes családjának bizonyos fontos tulajdonságaira mutat rá. Megadja a dekomponálhatóság szükséges és elégséges feltételét.

ON THE N-ARY DECOMPOSITION OF A RELATION IN THE CODD'S RELATION-MODEL

In the paper the concept of n-ary decomposition of a relation over a set of attributes is introduced. Some important properties of a full family of all decompositions are shown. The necessary and sufficient condition for a relation to be decomposable into n projections is given.

ОБ N-АРНОМ РАЗЛОЖЕНИИ СООТНОШЕНИЯ В РЕЛЯЦИОННОЙ МОДЕЛИ КОДДА

Ле Тиен Вуонг

Автор вводит понятие n-арного разложения и доказывает некоторые важные свойства полного семейства разложений. Он задает необходимое и достаточное условие разложимости.