数理解析研究所講究録
695巻 1989年 115-124

集合制約質問の記述方法および計算量

Representation and Complexity of Set Constraint Queries

岩井原瑞穂　　上林彌彦
九州大学工学部

1. まえがき

データベースに対する質問の答えは，条件を満たす組またはオブジェクトの集合であることが多い．これに対して，答えが条件を満たす集合，または組合せの集合であるような質問が考えられる．集合の個々の要素が条件を満たすだけではなく，集合全体として条件を満たすような質問である．このような条件を満たす集合を求める質問を，本稿では集合制約質問と呼ぶ．

例えば，次のような質問である．

【例1.1】 次のような組み合せを求める．

・各人の時間帯が重ならない作業の割り当て
・互いの距離がr以下の部品の集合
・年齢が5才離れている人のペア　□

データベース理論において，集合が陽に構成要素として扱われる例として，オブジェクト指向データベースや，非正規関係などが挙げられる[13]．このような場合，集合の構造はスキーマとして静的に固定されて扱われることが多く，集合制約質問のように，動的に集合を生成する機能はあまり考慮されていない．

正規形から非正規形への変換，すなわち集合を生成する操作はNEST/UNNEST[7]など，基本的に同じ値を持つ組を1つにまとめる操作のみで，組合せの条件を記述するには，表現力が不足する[12]．

論理データベースにおいては，再帰質問[2]を用いて推移的閉包や，親子関係での同世代の集合などを求めることが出来る．これらの質問は，最小不動点が存在することにより，答えは1つの集合となるが，組合せを求める問題では，解となる集合は複数存在するのが普通である．集合型の変数を導入することが自然であると考えられる．また，集合制約質問のクラスは推移的閉包を含む．

論理データベースへの集合型の導入に関する研究がある[1][9][10]．これらの文献では，述語論理の強力な記述力を用いた，大変表現力の高いデータベース言語が提案されている．本稿で考察する質問も，これらの言語で表すことができる．

これらの言語の処理系を実現する場合に，集合を扱うときに生じる特有の問

題がある．例えば，答えの数が指数的になる場合や，集合のネストの深さを任意にすると，少なくとも指数領域を必要とする質問が書けること[5]などである．

このような困難な質問も言語で記述できることは問題があり，除外されるように制限することが望ましいが，そのためには容易に求められる質問のクラスが明確にされねばならない．本稿では，集合制約質問で現実的な時間で求められる質問のクラスについて考察している．

集合制約質問の機能を備えるためには，集合全体を制約するような，条件を明確に記述できる手段が必要となる．本稿では，反射的・対称的である二項関係関係を与えられた条件とし，この二項関係の数学的性質より定義される集合を，質問の答えとする方法をとる．

例１．１においては，二項関係γ_rを用いて，答えＳは次のように表される．

$t_1 \ \gamma_r \ t_2 \quad iff \quad \| t_1[座標] - t_2[座標] \| \leq r$

$\forall t_1, t_2 \in S \quad ; \quad t_1 \gamma t_2$

以下２節では基本的事項を説明し，３節では集合制約質問の定義及び質問の分類を行う．そしてその部分クラスである，ＦＤ型集合制約質問について述べる．４節ではその処理法を検討する．質問の条件のＦＤが１つ及び２つの場合の多項式アルゴリズムを述べ，３つの場合はＮＰ完全になることを示す．また，データベースで保持されているＪＤを用いた処理の効率化を検討する．

２．　基本的事項

まず関係データベースについて説明する．

属性集合$X = \{A_1, A_2, \ldots, A_n\}$の各属性$A_i$は，定義域$D_i$を持つ．属性集合Ｘからなる関係Ｒ(Ｘ)は，$t : \{A_1, A_2, \ldots, A_n\} \rightarrow D_1 \times D_2 \times \cdots \times D_n$で表される写像ｔの有限集合であり，写像により各$A_i$は$D_i$の要素に写されるものとする．Ｘを関係スキーム，ｔをＲの組と呼ぶ．関係Ｒは列を属性，行を組とした，表の形で表現できる．本稿では，属性をＡ，Ｂ，Ｃ，･･･で表し，属性集合をＸ，Ｙ，･･･で表す．属性集合の連接は，その和集合を表す．

反射的・対称的な二項関係を同調関係といい，推移的な同調関係を同値関係という．Ａ上の同調関係γにおいて，Ａの部分集合Ｃで，Ｃの任意の２要素がγを満たすものを同調類という．

３．　集合制約質問

３．１　集合制約質問の定義

同調関係により定まる同調類は，同調関係による制約を受けた集合であるとみなすことができる．その集合は同調関係の定義を反映した性質を持つ．これ

を用いて，任意の要素が互いに似ているとか，同じである，近くにある，互いに独立である，交換可能である，などの集合を求めることができる．関係の組を１つの実体とし，組の属性値を実体の性質とみなし，質問で得たい組集合の特徴を，属性値で定義された２組間の同調関係で表すことにより，解の組集合を定めることができる．

【定義３．１】

関係Ｒ(Ｘ)の組集合を定義域とする，反射的かつ対称的な二項関係γが与えられたとき，γをＲ(Ｘ)上の集合制約質問と呼ぶ．質問の解は次の条件を満たす集合Ｓである．

$$\forall t_1, t_2 \in S \subset R(X);\quad t_1 \gamma t_2 \qquad \square$$

定義３．１では，Ｓが解ならばその部分集合も解となり，解の数は指数関数的になる．そのため，次の方法が考えられる．

①　極大集合を解とする．

しかしn個の点からなるグラフの 極大クリークの最大数は，$3^{n/3}$であるため，全ての解を出力することは適当でない．

②　極大集合の解を１つ出力する．

これは，組を解に１つづつ加えてゆくことにより，極大にでき，多項式時間で可能である．しかし出力される解は組の入力の順序により異なったものとなる．

③　最大集合の解を１つ出力する．

組合せを求める問題では，要素数最大の組合せが最適であることがよくある．しかし最大クリーク問題はＮＰ完全であり[11]，任意の質問に対する処理法を求めることは難しい．

３．２　　質問の分類

集合制約質問は，集合の定義に二項関係を用いているため，その定義の方法により，以下のような特徴づけができる．

①　基本論理式

述語論理において，真偽が与えられる最小単位の述語である．例えば２組の距離がｒ以下であるという，例１．１で定義した論理式γ_rなどである．他にも，２つの組のある属性の値が等しいという論理式等がある．

②　再帰的定義を含む場合

互いに連結している部品である，というようなことを表せる．

③ $\exists$，$\forall$などの限量子で修飾されている場合

④　ブール論理式

基本論理式がAND･OR･NOTで結合された，すなわちブール論理式で定義されたクラスがある．これはかなり単純なクラスと考えられる．

ブール論理式で定義された質問の例を示す．

【例３．１】

質問Q_3：図３．１に示す関係講義(科目,時限)は行われている全ての講義の科目と時限を表す．同じ科目の講義がいくつかの異なる時限で行なわれており，また同じ時限でいくつかの講義が行なわれているとする．

ここである学生が自分の時間割をつくることを考える．学生は同じ科目を２つ以上受講しないし，また同じ時限で２つ以上の科目は受講できない，そしてできるだけ多くの講義を受講できる時間割をつくりたい可能な講義の組み合せからなる講義の集合は次の条件を満たす．

① 各講義の科目はそれぞれ異なる．

② 同じ時限の講義は存在しない．

③ 集合の要素数は最大である．

科目	時限
英語	火２
英語	火２
物理	月２
物理	水１
数学	火２
数学	月２
数学	水１
化学	木１
化学	金２

図４．１

①と②より，関係・講義上の同調関係γ_cを定義する．

$$t_1 \ \gamma_c \ t_2 \quad \text{iff} \quad t_1 = t_2 \ \vee \ (t_1[\text{科目}] \neq t_2[\text{科目}] \ \wedge \ t_1[\text{時限}] \neq t_2[\text{時限}])$$

質問の答はγ_cの要素数最大の同調類である．□

質問Q_3は３．１節の③で述べた最大の集合を１つ求める質問である．

従来の関係データベースシステムでは，このような質問は応用プログラムによって，手続き的に記述されていた．

従属性はデータベース内の事実として，与えられている条件である．ＦＤ，ＭＶＤなどは，二項関係で表されることによって，集合制約質問と同じレベルで取り扱うことができる．例えば，関係Ｒ(ＸＹＺ)におけるＦＤ：Ｘ→Ｙは，以下のように表せる．

$$\forall t_1, t_2 \in R \ ; \quad t_1[X] = t_2[X] \quad \Rightarrow \quad t_1[Y] = t_2[Y] \qquad \square$$

”⇒”の部分を置き換えることにより，ＦＤは属性の等式による述語と，そのブール論理式で表せることがわかる[3].

【例３．２】 質問Q_3の条件は，以下に示すＦＤの集合に書き換えることができる．

F_1 ： 科目→時限

F_2 ： 時限→科目 □

3.3　FD型集合制約質問

ここでは，集合制約質問のある部分クラスについて考察する．

【定義3.2】　R(X)の組t_1, t_2について，属性A⊂Xについての基本論理式$t_1[A]=t_2[A]$を正リテラルと呼び，単にAで表す．正リテラルの否定を負リテラルと呼ぶ．正または負リテラルのAND･OR･NOTの結合で定義されたR上の二項関係をブール型二項関係と呼ぶ．　□

ブール型二項関係は必ずしも同調関係ではない．同調関係であるためには，主乗法標準形に変換したとき，各和項に少なくとも1つ正リテラルを含まねばならない．

ブール型二項関係の部分クラスを定義する．

【定義3.3】　ブール型二項関係γで表される集合制約質問について，γを主乗法標準形にしたとき，各和項に1つ正リテラルを含むものを，FD型集合制約質問と呼ぶ．　□

FD型同調関係の各和項$A_i \vee \neg B_{i1} \vee \neg B_{i2} \vee \cdots \vee \neg B_{ik(i)}$は，

FD：$B_{i1}B_{i2}\cdots B_{ik(i)} \rightarrow A_i$を表す．

FDは同調関係であるが，同値関係ではない．

ブール型二項関係質問は，論理関数の場合と同じ手法で簡単化できる[3]．

4.　FD型集合制約質問の処理法

本節では，FD型質問の解を求めるアルゴリズムについて考察する．条件のFD集合は，極小被覆に変換されているとする．FD集合の要素数により，場合分けをする．

4.1　FDが1つの場合

まず3.1節の①に対応する 極大の解を求めるアルゴリズムを示す．

【アルゴリズム1】

入力　：　関係R(組数m)，R上のFD　F_1

ステップ1：　関係RをFDの左辺が等しい組の集合$S_1,\ldots,S_p$に分ける．

ステップ2．　各S_iについてS_iをFDの右辺が等しい組の集合$S_{i1},\ldots,S_{iq(i)}$に分ける．

ステップ3：　$S = S_{1r(1)} \cup S_{2r(2)} \cup \cdots \cup S_{pr(p)}$　$(1 \leq r(i) \leq q(i))$とする．

Sは極大集合である．　□

アルゴリズム1において，解の個数がmの多項式に比例しない場合がある．そのため，FD型集合制約質問においても，全ての極大の解の個数は莫大なも

のとなり，すべて求めるのは適当でない．３．１節の② ③の，極大または最大の解を１つ求めるのが適当である．

４．２　　ＦＤが２つのときの最大の解

ＦＤが複数になると，１つの場合のように，単純なアルゴリズムを設計することは難しい．以下において，次の条件を満たすＦＤについて考察する．

【条件４．１】

Ｒ(Ｘ)上のＦＤ型集合制約質問の条件のＦＤ：Ｙ→Ｚについて，

Ｙ∪Ｚ＝Ｘとなる．　□

条件４．１を満たすＦＤの左辺は，Ｒ(Ｘ)のキーとなる．このため，解はＦＤの左辺の値が皆異なる組からなる集合である．これにより，質問Q_3のような，ある属性の値が皆違う，といった集合を表せる．

条件１を満たすＦＤにおいて，Ｒを左辺の値が等しい組集合に分け，その各集合をそれぞれ一つの超枝で覆った超グラフを考えると，最大の質問の解は，その超グラフの最大独立点集合に対応する．ここでは，各超枝からは１つしか点が選べないことを表す．

２つのＦＤの場合，超グラフに対し，超枝→点，点→有向枝の変換を行えば２部グラフが得られる．並行枝を除けば，最大解となる最大独立点集合は，２部グラフの最大マッチングに対応する．

【アルゴリズム２】

入力：関係Ｒ(Ｘ)及び条件４．１を満たすＦＤ　F_1，F_2

出力：F_1，F_2を満たすＲ(Ｘ)に含まれる最大の組集合

ステップ１：Ｒ(Ｘ)を，F_1の左辺の値が等しい集合 $S^1_1,\ldots,S^1_r$に分け，同様にF_2について$S^2_1,\ldots,S^2_s$にわける．すべてのS^1_i，S^2_iについて，それを超枝として持つ超グラフＨをつくる．

ステップ２：ある２つの超枝の共通部分に点が２つ以上含まれているならば，共通部分が１つの点になるようにする．

ステップ３：ＨのF_1の超枝の集合に点集合Ｕ，F_2の超枝の集合に点集合Ｖ，超枝の共通部分の点の集合に枝集合Ｅがそれぞれ１対１に対応する２部グラフＢ＝(Ｕ，Ｖ，Ｅ)を作る．

ステップ４：Ｂの最大マッチングＭを求め，それに対応するＨの点を独立点とする．□

【定理４．１】

条件４．１を満たす２つのＦＤからなる，Ｒ(組数ｎ)上の集合制約質問の最大の解は，$O(n^{1.5})$で求まる．

証明　アルゴリズム２のステップ１の超枝の個数はＯ(ｎ)であり，ステップ３まではＯ(ｎ log ｎ)でできる．ステップ４の最大マッチングは Ｏ($n^{1.5}$)のアルゴリズムが知られている[11]．　□

【例４．２】

質問Q_3に対して，アルゴリズム２を適用する．ステップ１の結果は 図４．３の超グラフとなり，ステップ３の２部グラフは図４．４となる．この最大マッチングを求めると，例えば枝(u_1, v_1)，(u_2, v_3)，(u_3, v_2)，(u_4, v_4)がマッチングとなる．各枝はそれぞれ組t_1, t_4, t_5, t_8に対応し，超グラフの独立点となる．図４．５が質問Q_3の解であり，F_1，F_2を満たしている．　□

	科目	時限	
t1	英語	火２	u 1
t2	英語	火２	
t3	物理	月２	u 2
t4	物理	水１	
t5	数学	火２	u 3
t6	数学	月２	
t7	数学	水１	
t8	化学	木１	u 4
t9	化学	金２	

図４．１　F_1の適用

	科目	時限	
t1	英語	月２	v 1
t6	数学	月２	
t3	物理	月２	
t2	英語	火２	v 2
t5	数学	火２	
t4	物理	水１	v 3
t7	数学	水１	
t8	化学	木１	v 4
t9	化学	金２	

図４．２　F_2の適用

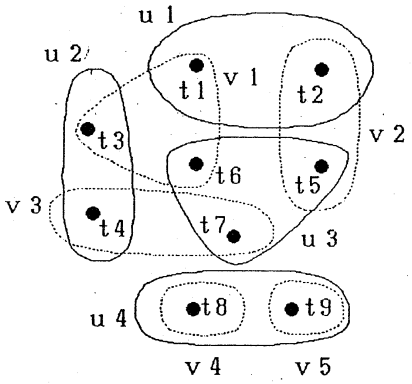

図４．３　超グラフH

	科目	時限
t1	英語	月２
t4	物理	水１
t5	数学	火２
t8	化学	木１

図４．５　最大の解

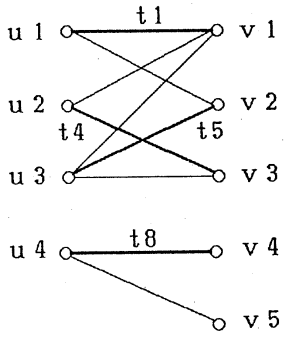

図４．４　２部グラフB

４．３　　ＦＤが３つ以上のときの最大の解

ここでは，ＦＤが３つの場合の集合制約質問の最大の解を求める問題が，ＮＰ完全であることを証明する．これより，３つ以上のＦＤの場合は，大変困難

になると考えられる.

３つのＦＤの問題はＮＰに含まれる. ＮＰ完全である３次元マッチング問題を集合制約質問に帰着する.

３次元マッチング問題[4]

インスタンス：集合Ｍ⊆Ａ×Ｂ×Ｃ. ここでＡ, Ｂ, Ｃは互いに素なｑ要素からなる集合である.

問題：Ｍ'⊆Ｍ, ｜Ｍ'｜＝ｑとなるマッチングＭ'が存在するか? ここでマッチングＭ'とはＭ'の各要素のＡ, Ｂ, Ｃの値が皆異なることを表す.

３次元マッチング問題は以下のようにして集合制約質問の解を求める問題に帰着できる.

３次元マッチング問題のインスタンスＭ⊆Ａ×Ｂ×Ｃに対応する３つの属性からなる関係Ｒ(ＡＢＣ)を作る. Ｒの組はＭの要素に対応する. 次の３つの(条件４.１を満たす) ＦＤからなるＲ上の集合制約質問をQ_2とする.

質問Q_2： Ａ→ＢＣ, Ｂ→ＣＡ, Ｃ→ＡＢ

要素数ｑのマッチングＭ'が 存在するかという問題は, 質問Q_2にｑ個の要素からなる解が存在するかという問題に帰着される. 以上により次の定理を得る.

【定理４.２】

３つの異なるＦＤによる集合制約質問は, ＮＰ完全である. □

４.４ ＪＤによる質問の変換

３.３節で述べたように, データベース管理のために, 常に保持されているＦＤ, ＭＶＤ(ＪＤ)などの従属性をＦＤ型集合制約質問の処理に利用することを検討する. ＦＤについては, 前述のように質問のＦＤ集合をデータベースのＦＤ集合で削減することが考えられる. ＪＤについては以下のような方法が考えられる.

関係Ｒ(ＸＹＺ)がＪＤ：*[ＸＹ, ＹＺ]を満たすとは, 以下の条件が成立するときである.

$$\forall t_1, t_2 \in R \;;\; t_1=(x_1\ y_1\ z_1),\ t_2=(x_1\ y_2\ z_2) \Rightarrow \exists t_3 \in R \text{ s.t. } t_3=(x_1\ y_1\ z_2)$$

直感的には, Ｘの１つの値にＹ, Ｚのそれぞれの値の集合が直積的に対応することを意味する. ２つの関係Ｒ(ＸＹ), Ｒ(ＸＺ)を結合, すなわちＸの値で２つの表を突き合わせて得られる, Ｒ(ＸＹＺ)は ＪＤ：*[ＸＹ, ＹＺ]を満たす. これより, 複数の関係を結合して得られた関係上での集合制約質問を処理する場合に, ＪＤが利用できる.

４.３節で用いた, ３つのＦＤからなる質問Q_2について考える. 質問の対象であるＲ(ＡＢＣ)上で, ＪＤ：*[ＡＢ, ＢＣ]が成り立つとする. このとき, Ｒ

（ＡＢＣ）は図４．６に 示すような非正規表現が可能である．属性Ａの点からＣの点へのパスの集合がＲ（ＡＢＣ）の組集合に対応する．図４．６の場合21個の組が存在する．属性Ａ，Ｃの部分には，重複した値が存在している．

質問Q_2により，Ａ，Ｂ，Ｃの値がすべて重複のない組集合を求める．

ＪＤ：*[ＡＢ，ＢＣ]が存在するとき，最大の解は２つのＦＤのときに２部グラフの最大マッチングを利用するのと同様に，最大流量問題を利用することができる．

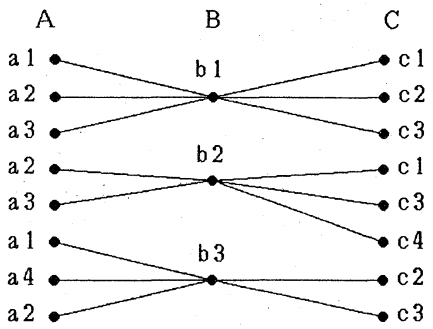

図４．６　ＪＤのある関係Ｒ（ＡＢＣ）

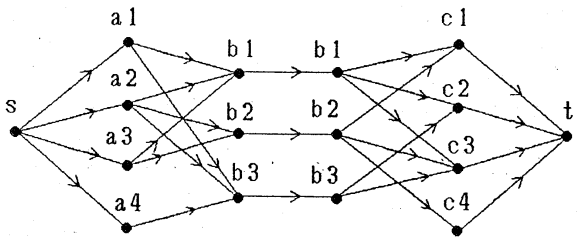

図４．７　最大流問題による解法

Ｒ（ＡＢＣ）を図４．７に示すネットワークＮ＝(s, t, Ｖ，Ａ，b)に変換する．sを流入口，tを流出口とし，全ての枝の容量を１とする．ＡとＢの間の枝は，分解された関係Ｒ（ＡＢ），同様にＢとＣの間の枝はＲ（ＢＣ）に対応する．

Ｎを流れる可能流について，sからtまで$a_i b_j c_k$の点を通る流れがあれば，組$(a_i b_j c_k)$を質問Q_2の解の組の１つとすることができる．なぜならば，Ａ，Ｂ，Ｃの各層において入る枝または出る枝が１本であり，また容量がすべて１であることより流れは常に１に飽和している．

そのためもし$a_i b_j c_k$を通る流れがあれば，a_i, b_j, c_k それぞれの点について，他にそこを通る流れは存在し得ないからである．

ｎ組の関係Ｒ（ＡＢＣ）から Ｏ（ｎ log ｎ）の手間でネットワークＮ＝(s, t, Ｖ，Ａ，b)を作ることができる．各節点に入る枝または出る枝が１本であり，全ての枝の容量が１である単純ネットワークの最大流を求める問題は，

$O(|V|^{1/2}|A|)$で求められる[11].

質問Q_2においては$|V|$，$|A|$ともにＯ（ｎ）であるため，$O(n^{1.5})$で最大の解を求めることができる．

以上のようなＪＤによる質問の変換が考えられる．次にＪＤの持つ性質より，この変換の適用可能性を検討する．

ＪＤには巡回と非巡回の重要なクラス分けが存在する．ＪＤの要素を超枝とする超グラフが巡回であるか調べることにより，ＪＤの巡回性を判定できる．

上述の質問Q_2におけるＪＤ：*[ＡＢ，ＢＣ]は，非巡回である．巡回であるＪ

Dの場合は質問処理に利用できるであろうか.

3次元マッチング問題は, インスタンスMが"pairwise consistent"に 制限されていてもNP完全である[4]. "pairwise consistent"とは, 任意のa, b, cについて (a, b, w)∈M, (a, x, c)∈M, (y, b, c)∈Mとなるw, x, yが存在するならば, (a, b, c)∈Mとなることである.

これはJD: *[AB, BC, CA]が成立していることと同値である. このJDは巡回である. このため巡回であるJDが存在したとしても, 質問処理は容易にならないと予想される.

5. まとめ

集合制約質問とその記述法について考察し, 部分クラスであるFD型集合制約質問の処理法について検討した. 他のクラスで効率的アルゴリズムが求められるものを見つけることが課題である.

参考文献

[1] Beeri, C., Naqvi, S., Ramakrishnan, R., et al, "Sets and Negation in a Logic Database Language(LDL1)", ACM PODS 1987, pp. 21-37.

[2] Chandra, A.K., "Theory of Database Queries", ACM PODS 1988, pp. 1-9.

[3] Delobel, C., Casey, R.G., "Decomposition of a Data Base and the Theory of Boolean Functions", IBM J. Res. Develop. 17, pp. 374-386, 1973.

[4] Garey, M.G., Johnson, D.S., "Computers and Intractability", FREEMAN, 1978.

[5] Hull, R., Su, J., "On the Expressive Power of Database Queries with Intermidate Types", ACM POD 1988, pp. 39-51.

[6] 岩井原, 上林, "二項関係によるデータベースの集合化質問", 昭和63年度電気関係学会九州支部連合大会論文集, pp. 930.

[7] Jaeschke, G., Schek, H-J. "Remarks on the algerbra of non-first-normal-form relations", ACM PODS, pp. 124-138, 1982.

[8] Kambayashi, Y., Tanaka, K., Yajima, S., "A Relational Data Language with Simplified Binary Relation Handling Capability", VLDB, pp. 338-350, 1977.

[9] Kuper, G.M., "Logic Programming With Sets", ACM PODS 1987, pp. 11-20.

[10] Kuper, G.M., "On the Expressive Power of Logic Programming Languages with Sets", ACM PODS 1988, pp. 11-14.

[11] Papadimitriou, C.H., Steiglitz, K., "Combinational Optimization", Prentice-Hall, 1982.

[12] Paredaens, J., "Possibilities and Limitations of Using Flat Operators in Nested Algebra Expressions", ACM PODS 1988, pp. 29-38.

[13] Scholl, M.H. & Schek, H.J. (Eds.), "Proc. Int. Workshop on Theory and Applications of Nested Relations and Complex Objects", Darmstadt, 1987.