データベース問合せ結果販売利益の分配に関する研究

鍋谷　優斗†　　吉川　正俊††

† 京都大学工学部情報学科　〒 606-8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科　〒 606-8501 京都府京都市左京区吉田本町
E-mail: †ynabetani@db.soc.i.kyoto-u.ac.jp, ††yoshikawa@i.kyoto-u.ac.jp

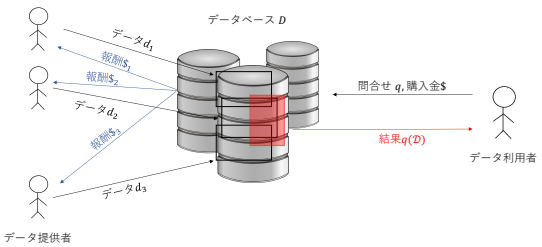

図 1　利益分配のイメージ

1. はじめに

日本は海外に比べてデータを活用したビジネス展開があまり進んでいない状況であり，情報銀行 [13] という事業が考案されている．情報銀行とは，個人とのデータ活用に関する契約等に基づき，PDS(Personal Data Store) と呼ばれる，他社保有データの集約を含め，個人が自らの意思で自らのデータを蓄積・管理するための仕組みを活用して，個人のデータを管理すると共に，個人の支持またはあらかじめ指定した条件に基づき個人に変わり妥当性を判断の上，データを第三者 (他の事業者) に提供する事業のことである．このような，個人情報を含めた多種多様で大量のデータの円滑な流通を実現するために個人の関与のもとでデータ流通・活用を進める仕組みが有効だと考えられている [10].

情報銀行を実現に近づけるため，パーソナルデータの価値付けを行うことは必要不可欠である．パーソナルデータの価値を考える研究は [8], [11] でも行われている．本研究においてもデータの価値付けにおける尺度及び価値付けの手法の開発を目指す．そもそも情報は他人に利用されて初めて価値があるとわかるものなので，実際に使われたデータから起源をたどり提供者に報酬を与える，という形が適していると考える．図 1 は，問合せを行う者が，その問合せ q とデータに対する報酬$をデータベース D に送り，データベース D は受け取ったに対する結果 $q(D)$ を問合せ者に返し，その結果のデータを提供した人物および結果のデータを得るために貢献した人物に報酬を分配する流れを示す．

データの起源を考えるという研究は [3] すでに多くのものがあり，Cheney らはサーベイ [3] で三つの起源 “Where-Provenance”, “How-Provenance”, “Why-Provenance” を紹介している．

そこで本研究では，2. で上記の研究について説明した後，4. でその研究を拡張し，各セルごとにデータの起源を示す直接来歴注釈と，そのデータがどのデータの貢献で存在しているかを示す間接来歴注釈を定義し，直接来歴注釈及び間接来歴注釈の与えられた関係データベースに対して四つの演算 (選択，射影，和，自然結合) の実行結果の関係における各セルの直接来歴注釈と間接来歴注釈の定義を与える．これを定義することで結果のデータを提供した人物および結果のデータを得るために貢献した人物を特定することができると考える．そして 4. で定義した演算を利用し，5. で問合せ結果に与えられた利益の分配方法についても考える．

2. 関連研究

文献 [3] では，データ起源について “Where-Provenance”, “How-Provenance”, “Why-Provenance” の三つの考え方を紹介している．

“Where-Provenance” は，ある関係に対する問合せの結果得られたデータのセルが元の関係のどのセルからきたものなのかを考えるもので，問合せ前の各セルそれぞれにラベル付けを行なっておくことで，問合せ結果の関係のセルが問合せ前の関係のどのセルから得られたものかを考えるものである．例えば図 2 のように二つの関係 R_1, R_2 を用意し，この二つの関係を自然結合した結果 $R_1 \bowtie R_2$ について “Where-Provenance” を考えると，まずはじめに自然結合を行う前の二つの関係のセルにそれぞれ $a_1, a_2, \ldots$ のようにラベルづけを行なっている．そして自然結合の結果の関係のセルが，元の二つの関係のどのセルから得られたものなのかを見て，そのセルのラベルを結果のセルに付与している．結果の関係の $(1,1)$ 成分のセル 1 は，R_1 の $(1,1)$ 成分 1 と R_2 の $(1,1)$ 成分 1 が結合してできたものなので，元の二つのセルのラベル a_1 と a_5 を結果のセルの 1 に付与している．“Where-Provenance” は等価な問合せに対して違う結果を返してしまう場合もあり，　[3] では等価な問合せの結果のラベルを全て含んだ結果を返すことで等価性を保つことができるということを紹介している．

次に，“How-Provenance” はある関係に対する問合せの結果得られたデータの組が，どのようにして得られたのかを考えてい

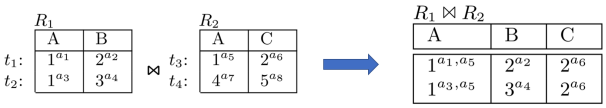

図 2　“Where-Provenance” におけるラベルづけ

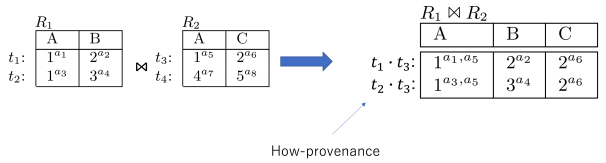

図 3　“How-Provenance” におけるタグづけ

る．結果の組がどのような演算の結果により得られたのかを和と積を用いて演算式で表現する．先ほどの関係 R_1, R_2 を自然結合した例で “How-Provenance” について考えると，図 3 で表される通り，まず二つの関係 R_1, R_2 の各組に t_1, t_2, t_3, t_4 のようにタグづけをして自然結合を行う．結果の関係の各組のタグは，その組の元となった組のタグ同士の積 (AND) で表現される．これはその二つのタグの両方が存在して初めて結果の関係が得られるということを意味している．“Why-Provenance” では，“How-Provenance” よりも簡略的に起源を考えるもので，演算式ではなく集合でそれを表現する．図 3 において一つ目の組の “ Why-Provenance” は $\{t_1, t_3\}$, 二つ目の組は $\{t_2, t_3\}$ となる．

Green ら [6] はデータ起源に関する種々の定義の関係についてまとめている．また，起源の情報自体を対象にする問合せ言語 [7] に関しても考えられている．

3. 準　　備

本報告書で利用する用語についてここで定義しておく．

ある関係 R における組 t の属性 A のセルを (t, A) で表し，関係 R が含む属性集合を $att(R)$ と表すものとする．

本報告書で対象とする問合せは，関係代数演算として選択 (selection)，射影 (projection)，自然結合 (natural join)，和 (union) を許す SPJU 問合せ，及び，SPJU 問合せのうち，選択の条件式として=（等号）と ∧(AND) のみを許す UCQ (Union of Conjunctive Query) である．

またある関係 R,S において，

条件式を c としてこの条件式を満たす組の集合を関係として返す演算を選択と言い，関係 R についてこの演算を行う場合，$\sigma_c R$ と表す．ここで，条件式 c については，最小の条件式を二項演算子 ∧(AND),∨(OR) で繋いだ条件式として捉える．最小の条件式とは，

（1） 属性と属性

（2） 属性と値

のどちらかの間に等号または不等号のみが入る条件式のことである．最小の条件式に ¬(NOT) が付いている場合はその条件式の余事象の条件式に置き換える．例えば，条件式 $c = (A = B) \wedge \neg(C = D)$ は，$\neg(C = D)$ を $(C = D)$ の余事象 $(C \neq D)$ に置き換えると，最小の条件式 $(A = B)$,$(C \neq D)$ を二項演算子 ∧ で繋いだ条件式と捉えることができる．

	A	**B**
t_1	1	2
t_2	2	3

図 4　R

	A
t_1	1
t_2	2

図 5　$\pi_A R$

B
2
3

図 6　$\pi_B(R_1 \bowtie R_2)$

関係 R の属性 A の上から射影する演算を $\pi_A R$ と表す．関係 R, S の和演算を $R \cup S$ と表す．関係 R, S の自然結合演算を $R \bowtie S$ と表す．

4. 直接来歴注釈，間接来歴注釈の定義と計算

本研究では，問合せ結果のデータの起源を辿り，元となっているセルに価値を与えたい．そのため Green ら [5] の，和積の形で結果を算出する “How-Provenance” の考えをセル単位で適用する．組単位ではなくセル単位で考える理由としては，組ごとでは最終結果に残っておらず，演算にも使用されていないデータにまで価値を与える可能性があるからである．また，セル単位で考えると組単位で考えるよりもより詳細にデータに価値を与えることができる．例えば，図 4 のように組にタグ t_1, t_2 が与えられた関係 R の属性 A の上に射影演算を行った結果は図 5 のようになるが，組ごとに価値を与えると，t_1, t_2 が結果に残っていない属性 B の値 2,3 も含んでいるため，これらの値にも価値を与えることになる．

ここで，問合せによって得られたそのセルが元のどのセルから得られたものなのかを考えるために，問合せによって演算が行われる前のデータのセルにそれぞれデータの起源を示す直接来歴注釈を与える．さらに，問合せによって得られたそのセルは，どのセルの貢献によって得られたものなのかも考えるため，間接来歴注釈も与える．間接来歴注釈を考える理由としては，起源そのものを考えるだけでは，そのデータを得るために貢献したデータには価値を与えることができない可能性があるからである．例えば，図 2 の結果の関係の属性 B の上から射影する演算を行なった場合，図 6 のような結果が得られるが，この二つのセルの値 2,3 が存在するのは図 2 で関係 R_1 の属性 A における二つの値 1 と，関係 R_2 の属性 A における値 1 が同じ値を取っており，自然結合することができたからである．そのため問合せ結果には残っていないが関係 R_1 の属性 A における二つの値 1 と，関係 R_2 の属性 A における値 1 にも報酬を与えるべきである．

つまり，演算において注目しているセルが存在しているのはどのデータの貢献によるのかを知るために間接来歴注釈を考えている．直接来歴注釈，間接来歴注釈は簡単にはラベルのようなものだが，演算にも利用可能できるより広い概念のものとしている．この二つを問合せ前の関係の各セルに追加し，問合せによる演算の結果得られたデータの各セルの直接来歴注釈，間接来歴注釈はどのようになるかを考える．Buneman ら [1] もデータを木構造として捉え，独自の問合せ言語を用いて広い範囲で起源を考える方法を示しているが，本研究では関係代数演算を前提とする．選択，射影，和，自然結合の四つの演算を用いた

name	based_in	phone
BayTours (s_{11}, l_{11})	San Francisco (s_{12}, l_{12})	415-1200 (s_{13}, l_{13})
HarborCruz (s_{21}, l_{21})	Santa Cruz (s_{22}, l_{22})	831-3000 (s_{23}, l_{23})

図 7　*Agencies*

name	destination	type
BayTours (s'_{11}, l'_{11})	San Francisco (s'_{12}, l'_{12})	cable car (s'_{13}, l'_{13})
BayTours (s'_{21}, l'_{21})	Santa Cruz (s'_{22}, l'_{22})	bus (s'_{23}, l'_{23})
BayTours (s'_{31}, l'_{31})	Santa Cruz (s'_{32}, l'_{32})	boat (s'_{33}, l'_{33})
BayTours (s'_{41}, l'_{41})	Monterey (s'_{42}, l'_{42})	boat (s'_{43}, l'_{43})

図 8　*ExternalTours*

name	destination	type
BayTours (s''_{11}, l''_{11})	San Francisco (s''_{12}, l''_{12})	cable car (s''_{13}, l''_{13})
HarborCruz (s''_{21}, l''_{21})	Santa Cruz (s''_{22}, l''_{22})	bus (s''_{23}, l''_{23})

図 9　*OriExternalTours*

問合せを行った場合に，結果の関係の各セルの直接来歴注釈，間接来歴注釈成分の計算方法を以下で定義する．
セル (t, A) の直接来歴注釈，間接来歴注釈をそれぞれ $source(t, A)$，$lineage(t, A)$ で表すものとする．ある間接来歴注釈 l に対しては，べき等律，すなわち

$$l \cdot l = l$$

が成り立つものとする．
また，具体例として図 7,8,9 のデータベースを用いることとする． 図 7,8,9 の各セルの成分は第一成分が直接来歴注釈，第二成分が間接来歴注釈を表す．例では予め上記の直接来歴注釈，間接来歴注釈を持っているものとする．
本研究の目標は問合せに基づいて利益を分配することである．そのために，等価であるが問合せ式が異なる問合せの扱い方には注意が必要がある．

（1） 等価であるが問合せ式が異なる問合せに対しては同じように利益を分配するべきだという考えの場合

（2） 等価であるが問合せ式が異なる問合せに対する利益の分配結果が異なっても良いという考えの場合

に分けて考え方を提案する．
まず等価な問合せに対しては同じように利益を分配するべきだという考えの場合について考える．ここで，選択の選択条件において =(等号) と ∧ (AND) のみを許す UCQ 問合せ (conjunctive query) では，等価な問合せを，等価性を保ったまま最小化した問合せが少なくとも一つ存在するということが証明されている [2], [12, 6 章 2 節]．この場合，問合せを最小化したものに対して直接来歴注釈，間接来歴注釈計算を行えば，必ず問合せに対する直接来歴注釈，間接来歴注釈の値は一意に定まる．例えば，図 10 を見ると，二つの等価な問合せ (問合せ 1 と問合せ 2) に対してまず最小化を行い，その後最小化した問合せに

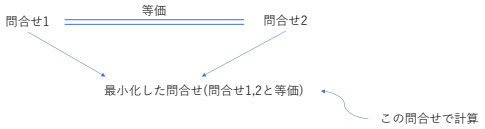

図 10　等価な問合せへの対処

表 1　等価な問合せへの対処

問合せクラス / 等価な問合せ	UCQ 問合せ	SPJU 問合せ
等価ならば利益分配も同じ	問合せの最小化	-
等価でも利益分配が異なってよい	問合せによって直接来歴注釈，間接来歴注釈は異なる	

対して直接来歴注釈，間接来歴注釈計算を行えばどちらの問合せに対しても同じ値が得られる．
文献 [3] の第 4 章 “Where-Provenance” では等価な問合せに対して，結果のセルにつくラベルを全てつけておくことで等価性を保つ方法を紹介しているが，文献 [4] ではそれを否定し，[3] の第 2 章 “Why-Provenance” で述べられている，等価な問合せに対して最小化を行う方法を肯定している．そのため本報告書においても等価な問合せに対しては最小化を行うことで対処する．よって，選択条件に =(等号) と ∧ (AND) のみを許す UCQ 問合せの場合はその問合せを最小化した上で，最小化した問合せに対して直接来歴注釈，間接来歴注釈計算を行い，それに基づいて価値づけを行うようにする．選択条件に不等号や ∨ (OR), ¬(NOT) を含む SPJU 問合せの場合はこの考え方は適用できない．
次に，等価な問合せに対する利益の分配結果は異なっても良いという考えの場合については，等価であっても問合せごとに直接来歴注釈，間接来歴注釈が異なっても良いとし，計算を行うこととする．
以上のことを表にまとめると表 1 のようになる．4.1 節で SPJU 問合せにおける直接来歴注釈，間接来歴注釈計算方法を定義し，4.2 節で UCQ 問合せにおける直接来歴注釈，間接来歴注釈計算方法を定義する．

4.1　SPJU 問合せの場合

4.1.1　選　　択

関係 R に対してある選択条件 c による選択を行った結果の関係 $\sigma_c R$ の中のあるセル (t_a, A) の直接来歴注釈，間接来歴注釈を考える．ここで，t_a, A は関係 R のある組，および属性を表す．
ここで，選択条件 c は最小の条件式集合と ∧ (AND), ∨ (OR), ¬(NOT) の 3 つの記号で構成されているものとする．最小の条件式に ¬ が付いている場合は 3. の ¬ に対する操作をあらかじめ行っておく．
演算の結果の関係のセルが元の関係のセル (t, A) から得られたものとする．この時，セル (t_a, A) の直接来歴注釈，間接来歴注釈は，ある間接来歴注釈 l_{sub} を用いて以下のように書ける．

$$source(t_a, A) = source(t, A)$$

$$lineage(t_a, A) = lineage(t, A) \cdot l_{sub}$$

ここで，l_{sub} は選択条件 c によって決定される．

name	destination	type
BayTours $(s'_{11}, l'_{11} \cdot (l'_{12} + l'_{13}))$	San Francisco $(s'_{12}, l'_{12} \cdot (l'_{12} + l'_{13}))$	cable car $(s'_{13}, l'_{13} \cdot (l'_{12} + l'_{13}))$
BayTours $(s'_{21}, l'_{21} \cdot l'_{23})$	Santa Cruz $(s'_{22}, l'_{22} \cdot l'_{23})$	bus (s'_{23}, l'_{23})
BayTours $(s'_{41}, l'_{41} \cdot l'_{42})$	Monterey (s'_{42}, l'_{42})	boat $(s'_{43}, l'_{42} \cdot l'_{43})$

図 11　$\sigma_c(ExternalTours)$

選択条件 c が，その全体を二分する記号によって，$c = C_1 \wedge C_2$ または $c = C_1 \vee C_2$ で構成されている場合，l_{sub} は C_1 及び C_2 の間接来歴注釈成分をそれぞれ l_{C_1}, l_{C_2} とすると，以下のように計算される．

（1）　$c = C_1 \wedge C_2$ の場合

$$l_{sub} = l_{C_1} \cdot l_{C_2}$$

（2）　$c = C_1 \vee C_2$ の場合

$$l_{sub} = l_{C_1} + l_{C_2}$$

条件式 C_1, C_2 の間接来歴注釈成分 l_{C_1}, l_{C_2} に対して上記の操作を，条件式が最小の条件式になるまで再帰的に利用する．その後，各最小の条件式についての間接来歴注釈成分は以下で求める．ある最小の条件式を c_{sub} とし，c_{sub} を満たす属性集合を B とすると，それぞれの間接来歴注釈成分 l_c は以下である．

$$l_c = \prod_{b \in B} lineage(t, b)$$

以上の手順によって得られた l_{sub} を代入することで求めたい間接来歴注釈が得られる．なお ¬(NOT) に関しては，ド・モルガンの法則によって最小の条件式の否定にできるため，上記の手順で計算可能である．

［例 1］　条件 $c : \neg((destination =' SantaCruz') \wedge (type =' boat'))$ とした時の $\sigma_c(ExternalTours)$ による結果の成分を考えると，図 11 のようになる．

例えば図 11 の (1,1) 成分に注目すると，この $BayTours$ の直接来歴注釈，間接来歴注釈成分の計算は以下のようにしてなされる．

直接来歴注釈に関しては，そのセルの元のセルの直接来歴注釈がそのまま注目しているセルの直接来歴注釈になるため，s'_{11} となる．

次に間接来歴注釈に関しては，そのセルの元のセルの間接来歴注釈に l_{sub} をかけたものがそのセルの間接来歴注釈になるため，l_{sub} を計算する．まず条件式の外側に ¬ 記号があるので，最小の条件式の not になるまでド・モルガンの法則を用いると，

$\neg((destination =' SantaCruz') \wedge (type =' boat'))$

$\equiv$

$(\neg(destination =' SantaCruz') \vee \neg(type =' boat'))$ $(\equiv$ $((destination \neq' SantaCruz') \vee (type \neq' boat')))$

とできる．次にこの条件式は全体を二分する記号 ∨ によって構成されているため，

destination
San Francisco (s'_{12}, l'_{12})
Santa Cruz $(s'_{22} + s'_{32}, l'_{22} + l'_{32})$
Monterey (s'_{42}, l'_{42})

図 12　$\pi_{destination}(ExternalTours)$

$$C_1 = \neg(destination =' SantaCruz')$$

$$C_2 = \neg(type =' boat')$$

に分けることができ，C_1 及び C_2 の間接来歴注釈成分をそれぞれ l_{C_1}, l_{C_2} とすると，l_{sub} は以下のようになる．

$$l_{sub} = l_{C_1} + l_{C_2}$$

条件式 C_1, C_2 は最小の条件式のため，l_{C_1}, l_{C_2} がそれぞれ求まる．(1,1) 成分が属する組 t は C_1 を満たしており，C_1 に現れる属性集合 $B = \{destination\}$ であるため，l_{C_1} は以下のようになる．

$$l_{C_1} = l'_{12}$$

C_2 に関して，(1,1) 成分が属する組 t は C_2 を満たしており，C_2 に現れる属性集合 $B' = \{type\}$ なので，l_{C_2} は以下のようになる．

$$l_{C_2} = l'_{13}$$

以上より，$l_{sub} = l'_{12} + l'_{13}$ となり，求める (1,1) 成分の間接来歴注釈成分は，

$$l_{11} \cdot (l'_{12} + l'_{13})$$

となる．

4.1.2　射　　影

関係 R を属性集合 X 上に射影した結果の関係 $\pi_X R$ の中のあるセル (t_a, A) の直接来歴注釈，間接来歴注釈を考える．これらは，それぞれ，元の関係 R の組で X 上に射影すると t_a になるものを t とすると t の直接来歴注釈，間接来歴注釈の和とする．すなわち，

$$source(t_a, A) = \sum_{t \in R, \pi_X t = t_a} source(t, A)$$

$$lineage(t_a, A) = \sum_{t \in R, \pi_X t = t_a} lineage(t, A)$$

［例 2］　図 8 の $ExternalTours$ を属性 $destination$ の上に射影した場合

$\pi_{destination}(ExternalTours)$ を考えると，結果は図 12 のようになる．射影は重複削除を行うため，同じ値をもつセルは一つになる．上記の図においては元の $ExternalTours$ では $SantaCruz$ が二つ存在するため，射影後は一つになっている．直接来歴注釈，間接来歴注釈成分はそれぞれ，元の二つのセルの成分の和で表され，$(s'_{22} + s'_{32}, l'_{22} + l'_{32})$ となる．$SantaCruz$

name	destination	type
BayTours $(s'_{11}+s''_{11}, l'_{11}+l''_{11})$	San Francisco $(s'_{12}+s''_{12}, l'_{12}+l''_{12})$	cable car $(s'_{13}+s''_{13}, l'_{13}+l''_{13})$
BayTours (s'_{21}, l'_{21})	Santa Cruz (s'_{22}, l'_{22})	bus (s'_{23}, l'_{23})
BayTours (s'_{31}, l'_{31})	Santa Cruz (s'_{32}, l'_{32})	boat (s'_{33}, l'_{33})
BayTours (s'_{41}, l'_{41})	Monterey (s'_{42}, l'_{42})	boat (s'_{43}, l'_{43})
HarborCruz (s''_{21}, l''_{21})	Santa Cruz (s''_{22}, l''_{22})	bus (s''_{23}, l''_{23})

図 13　$(ExternalTours) \cup (OriExternalTours)$

以外の成分はもともと一つしか存在しないため，元のセルの直接来歴注釈，間接来歴注釈成分そのままとなる．

4.1.3　和

関係 R, S の和演算による結果の関係 $R \cup S$ の中のあるセル (t_a, A) の直接来歴注釈，間接来歴注釈を考える．結果の関係の組 t_a が，それぞれもとの関係 R と関係 S に存在する同一の組 t_R, t_S に対応する場合，t_a のすべての属性 A に対し，

$$source(t_a, A) = source(t_R, A) + source(t_S, A)$$

$$lineage(t_a, A) = lineage(t_R, A) + lineage(t_S, A)$$

結果の関係の組 t_a がもとの関係 R の組 t_R のみに対応する場合，t_a のすべての属性 A に対し，

$$source(t_a, A) = source(t_R, A)$$

$$lineage(t_a, A) = lineage(t_R, A)$$

結果の関係の組 t_a がもとの関係 S の組 t_S のみに対応する場合，t_a のすべての属性 A に対し，

$$source(t_a, A) = source(t_S, A)$$

$$lineage(t_a, A) = lineage(t_S, A)$$

［例 3］ 図 8,9 の $ExternalTours$ と $OriExternalTours$ の和をとった場合
$(ExternalTours) \cup (OriExternalTours)$ を考える．結果は図 13 のようになる．図 13 において，1 つ目の組が関係 $ExternalTours$ と関係 $OriExternalTours$ の同じ組が合わさってできた組になっているので，その組の各セルの直接来歴注釈，間接来歴注釈成分は，それぞれ元の二つのセルの成分の和で表され，例えば (1,1) 成分 (属性 *name* の BayTours) に注目すると，その成分は $(s'_{11}+s''_{11}, l'_{11}+l''_{11})$ で表される．その他の組の成分に関しては，それぞれ他方の関係の組と異なっているため，元のセルの直接来歴注釈，間接来歴注釈成分になる．

4.1.4　自 然 結 合

関係 R, S の自然結合による結果の関係 $R \bowtie S$ の中のあるセル (t_a, A) の直接来歴注釈，間接来歴注釈を考える．結果の関係の組 t_a が元の関係 R と関係 S の組 t_R, t_S の自然結合に対応する場合，t_a の属性のうち

（1）　$A_J \in att(R) \cap att(S)$

name	based_in	phone	destination	type
BayTours $(s_{11} \cdot s'_{11}, l_{11} \cdot l'_{11})$	San Francisco $(s_{12}, l_{12} \cdot l_{11} \cdot l'_{11})$	415-1200 $(s_{13}, l_{13} \cdot l_{11} \cdot l'_{11})$	San Francisco $(s'_{12}, l'_{12} \cdot l_{11} \cdot l'_{11})$	cable car $(s'_{13}, l'_{13} \cdot l_{11} \cdot l'_{11})$
BayTours $(s_{11} \cdot s'_{21}, l_{11} \cdot l'_{21})$	San Francisco $(s_{12}, l_{12} \cdot l_{11} \cdot l'_{21})$	415-1200 $(s_{13}, l_{13} \cdot l_{11} \cdot l'_{21})$	Santa Cruz $(s'_{22}, l'_{22} \cdot l_{11} \cdot l'_{21})$	bus $(s'_{23}, l'_{23} \cdot l_{11} \cdot l'_{21})$
BayTours $(s_{11} \cdot s'_{31}, l_{11} \cdot l'_{31})$	San Francisco $(s_{12}, l_{12} \cdot l_{11} \cdot l'_{31})$	415-1200 $(s_{13}, l_{13} \cdot l_{11} \cdot l'_{31})$	Santa Cruz $(s'_{32}, l'_{32} \cdot l_{11} \cdot l'_{31})$	boat $(s'_{33}, l'_{33} \cdot l_{11} \cdot l'_{31})$
BayTours $(s_{11} \cdot s'_{41}, l_{11} \cdot l'_{41})$	San Francisco $(s_{12}, l_{12} \cdot l_{11} \cdot l'_{41})$	415-1200 $(s_{13}, l_{13} \cdot l_{11} \cdot l'_{41})$	Monterey $(s'_{42}, l'_{42} \cdot l_{11} \cdot l'_{41})$	boat $(s'_{43}, l'_{43} \cdot l_{11} \cdot l'_{41})$

図 14　$(Agencies) \bowtie (ExternalTours)$

（2）　$A_R \in att(R) - att(A_J)$

（3）　$A_S \in att(S) - att(A_J)$

とすると、
直接来歴注釈：

$$source(t_a, A_J) = source(t_R, A_J) \cdot source(t_S, A_J)$$

$$source(t_a, A_R) = source(t_R, A_R)$$

$$source(t_a, A_S) = source(t_S, A_S)$$

間接来歴注釈：

$$lineage(t_a, A_J) = \prod_{A_J} (lineage(t_R, A_J) \cdot lineage(t_S, A_J))$$

$$lineage(t_a, A_R) = lineage(t_R, A_R) \cdot \prod_{A_J} (lineage(t_R, A_J) \cdot lineage(t_S, A_J))$$

$$lineage(t_a, A_S) = lineage(t_R, A_S) \cdot \prod_{A_J} (lineage(t_R, A_J) \cdot lineage(t_S, A_J))$$

［例 4］ 関係 $Agencies$ と関係 $ExternalTours$ を自然結合した場合 $(Agencies) \bowtie (ExternalTours)$ を考える．結果は図 14 のようになる．具体例において，それぞれの属性 *name* が両方 BayTours を持っているため，これで結合がなされる．自然結合も同じ属性の同じ値は重複削除がなされて一つになるため，結合の元となった BayTours の直接来歴注釈成分両方の積が結果の関係のセルの直接来歴注釈成分となる．間接来歴注釈成分は，結合の元となった BayTours の間接来歴注釈成分それぞれの積が，その組のセルそれぞれの元々の間接来歴注釈成分にかけたものとなる．
例えば，図 14 の一つ目の組に注目すると，(1,1) 成分は結合の元となった二つの BayTours が合わさってできたセルなので直接来歴注釈，間接来歴注釈成分はそれぞれ元の二つのセルの成分の積となり，$(s_{11} \cdot s'_{11}, l_{11} \cdot l'_{11})$ となる．(1,2) 成分から (1,5) 成分については，それぞれ元の二つの関係の共通でない属性のセルなので，直接来歴注釈成分は元のセルの直接来歴注釈成分のままとなる．間接来歴注釈に関しては，結合の元となった (1,1) 成分の間接来歴注釈が元のセルの間接来歴注釈にかけられる．例えば (1,2) 成分は，元の間接来歴注釈成分である l_{12} に，結合の元となった (1,1) 成分の間接来歴注釈成分である $l_{11} \cdot l'_{11}$ がかけられ，結果として $(s_{12}, l_{12} \cdot l_{11} \cdot l'_{11})$ が得られる．

Result of Q:

destination	phone	
San Francisco	415-1200	$t_1 \cdot (t_1 + t_3)$
Santa Cruz	831-3000	t_2^2
Santa Cruz	415-1200	$t_1 \cdot (t_4 + t_5)$
Monterey	415-1200	$t_1 \cdot t_6$

図 16　図 15 の問合せによる結果

4.2　UCQ 問合せの場合

選択条件に =(等号)，∧(AND) のみを許す論理積問合せ (conjunctive query) については，選択のみ計算が先ほどよりも容易になる．他の三つの演算については 4.1 節と同じである．よって選択の直接来歴注釈，間接来歴注釈の計算方法のみ新たに定義する．一般的には先ほどと同様に，セル (t_a, A) の直接来歴注釈，間接来歴注釈は，ある間接来歴注釈 l_{sub} を用いて以下のように書ける．

$$source(t_a, A) = source(t, A)$$

$$lineage(t_a, A) = lineage(t, A) \cdot l_{sub}$$

ここで，先ほどと同様に l_{sub} は選択条件 c によって決定されるが，計算方法は先ほどよりも容易である．具体的には，選択条件 c に現れる最小の条件式 (等式で表されている) に含まれる属性の集合を B とすると，l_{sub} は以下のように計算できる．

$$l_{sub} = \prod_{b \in B} lineage(t, b)$$

以上の手順によって得られた l_{sub} を代入することで求めたい間接来歴注釈が得られる．

4.3　具　体　例

文献 [3] の第 3 章 “How-Provenance” で使用されている例では，図 7,8 に対して以下の問合せを実行している．

```
Q:
SELECT    e.destination, a.phone
FROM      Agencies a,
          (SELECT name,
               based_in AS destination
          FROM Agencies a
          UNION
          SELECT name, destination
          FROM ExternalTours ) e
WHERE     a.name = e.name
```

図 15　[3] 第 3 章 “How-Provenance” p418 の例

この問合せを行った結果が図 16 のようになる．この問合せを，先ほど定義した手法で実行してセルごとに成分を計算する．SQL で表されているこの問合せを関係代数で表現すると，

$$\pi_{destination,phone}(\\ Agencies \\ \bowtie \\ ((\delta_{based_in \to destination}(\pi_{name,based_in} Agencies)) \\ \cup \\ (\pi_{name,destination} ExternalTours)))$$

name	destination
BayTours (s_{11}, l_{11})	San Francisco (s_{12}, l_{12})
HarborCruz (s_{21}, l_{21})	Santa Cruz (s_{22}, l_{22})

図 17　Q_1 の結果の来歴注釈

name	destination
BayTours (s'_{11}, l'_{11})	San Francisco (s'_{12}, l'_{12})
BayTours $(s'_{21} + s'_{31}, l'_{21} + l'_{31})$	Santa Cruz $(s'_{22} + s'_{32}, l'_{22} + l'_{32})$
BayTours (s'_{41}, l'_{41})	Monterey (s'_{42}, l'_{42})

図 18　Q_2 の結果の来歴注釈

name	destination
BayTours $(s_{11} + s'_{11}, l_{11} + l'_{11})$	San Francisco $(s_{12} + s'_{12}, l_{12} + l'_{12})$
HarborCruz (s_{21}, l_{21})	Santa Cruz (s_{22}, l_{22})
BayTours $(s'_{21} + s'_{31}, l'_{21} + l'_{31})$	Santa Cruz $(s'_{22} + s'_{32}, l'_{22} + l'_{32})$
BayTours (s'_{41}, l'_{41})	Monterey (s'_{42}, l'_{42})

図 19　$Q_1 \cup Q_2$ の結果の来歴注釈

name	based_in	phone	destination
BayTours $(s_{11} \cdot (s_{11} + s'_{11}), l_{11} \cdot (l_{11} + l'_{11}))$	San Francisco $(s_{12}, l_{12} \cdot l_{11} \cdot (l_{11} + l'_{11}))$	415-1200 $(s_{13}, l_{11} \cdot l_{13} \cdot (l_{11} + l'_{11}))$	San Francisco $(s_{12} + s'_{12}, l_{11} \cdot (l_{11} + l'_{11}) \cdot (l_{12} + l'_{12}))$
HarborCruz $(s_{21} \cdot s_{21}, l_{21})$	Santa Cruz $(s_{22}, l_{21} \cdot l_{22})$	831-3000 $(s_{23}, l_{21} \cdot l_{23})$	Santa Cruz $(s_{22}, l_{21} \cdot l_{22})$
BayTours $(s_{11} \cdot (s'_{21} + s'_{31}), l_{11} \cdot (l'_{21} + l'_{31}))$	San Francisco $(s_{12}, l_{11} \cdot l_{12} \cdot (l'_{21} + l'_{31}))$	415-1200 $(s_{13}, l_{11} \cdot l_{13} \cdot (l'_{21} + l'_{31}))$	Santa Cruz $(s'_{22} + s'_{32}, l_{11} \cdot (l'_{22} + l'_{32}) \cdot (l'_{21} + l'_{31}))$
BayTours $(s_{11} \cdot s'_{41}, l_{11} \cdot l'_{41})$	San Francisco $(s_{12}, l_{11} \cdot l_{12} \cdot l'_{41})$	415-1200 $(s_{13}, l_{11} \cdot l_{13} \cdot l'_{41})$	Monterey $(s'_{42}, l_{11} \cdot l'_{41} \cdot l'_{42})$

図 20　$Agencies \bowtie (Q_1 \cup Q_2)$ の結果の来歴注釈

destination	phone
San Francisco $(s_{12} + s'_{12}, l_{11} \cdot (l_{11} + l'_{11}) \cdot (l_{12} + l'_{12}))$	415-1200 $(s_{13}, l_{11} \cdot l_{13} \cdot (l_{11} + l'_{11}))$
Santa Cruz $(s_{22}, l_{21} \cdot l_{22})$	831-3000 $(s_{23}, l_{21} \cdot l_{23})$
Santa Cruz $(s'_{22} + s'_{32}, l_{11} \cdot (l'_{22} + l'_{32}) \cdot (l'_{21} + l'_{31}))$	415-1200 $(s_{13}, l_{11} \cdot l_{13} \cdot (l'_{21} + l'_{31}))$
Monterey $(s'_{42}, l_{11} \cdot l'_{41} \cdot l'_{42})$	415-1200 $(s_{13}, l_{11} \cdot l_{13} \cdot l'_{41})$

図 21　Q の結果の来歴注釈

となる．まず，

$Q_1 \equiv (\delta_{based_in \to destination}(\pi_{name,based_in} Agencies))$

と

$Q_2 \equiv (\pi_{name,destination} ExternalTours))$

の部分について成分を計算すると図 17,18 のようになる． 次に，$Q_1 \cup Q_2$ について成分を計算すると図 19 のようになる．この関係と関係 $Agencies$ を自然結合すると図 20 のようになる．最後にこの関係を $destination, phone$ の上に射影したものが求める関係である．結果は図 21 のようになる．

5.　問合せ結果に貢献したセルへの利益分配

この章では 4. において定義した直接来歴注釈，間接来歴注釈を利用して，得られた結果に与えられた利益をデータ提供者にどのような割合で分配するのかを考える．データの価値づけは [9] のように情報量に基づいて価値を考える研究もなされて

いる．
予め以下の定義を行っておく．
得られたデータに W の価値があるとする．選択，射影，和，自然結合の四つの演算に対し，データの価値付けを以下のように定義する．この時演算の結果得られた関係のセルに，順に価値 $w_i, (i = 1, 2, ...)$ を割り当てるものとする．ここで，w_i は，

$$W = \sum_i w_i$$

が成り立つように決定する．
さらに，直接来歴注釈，間接来歴注釈成分にそれぞれ価値 ω_s, ω_l を

$$w_i = \omega_s + \omega_l$$

となるように与えるものとする．
また，あるセル c の価値を $val(c)$ と表す．この関数は直接来歴注釈 s, または間接来歴注釈成分 l を引数としてとる場合，それぞれを提供したデータに $val(s)$ または $val(l)$ の価値があると考える．
関数 $add(s, v)$ は $val(s)$ に v を加算したものを新たに $val(s)$ とすることを意味する ($val(s)$ を更新する)．
この定義を元に第 3 章と同様に，選択，射影，和，自然結合の四つの演算をそれぞれ一回だけ行った場合の価値づけの方法を以下で定義する．

5.1 射影，和

価値 w が与えられたセル c の直接来歴注釈，間接来歴注釈成分を (s, l) とすると，射影, または和において，n を自然数として

$$s = s_1 + s_2 + ... + s_n$$

$$l = l_1 + l_2 + ... + l_n$$

と表すことができる．このとき，

$$add(s_i, \frac{\omega_s}{n})$$

$$add(l_i, \frac{\omega_l}{n})$$

具体例として，図 12,13 にそれぞれ価値 100 を与えたとする．この時データ利用者はどのセルにどれだけ報酬を与えるかを指定することができる (もし指定がなければ得られたデータの全セルに均等に価値 100 を分配する)．また，ここでは直接来歴注釈，間接来歴注釈に

$$\omega_s : \omega_l = 6 : 4$$

の比で価値を与えるように指定しておく．
セルに対して以下のように価値が与えられたとする． 図 22 において，一行一列目の “San Francisco” には価値 10 が与えられている．この時，

$$\omega_s : \omega_l = 6 : 4$$

より，

$$add(s'_{12}, 6)$$

destination
San Francisco (s'_{12}, l'_{12}) 10
Santa Cruz $(s'_{22} + s'_{32}, l'_{22} + l'_{32})$ 60
Monterey (s'_{42}, l'_{42}) 30

図 22 $\pi_{destination}(ExternalTours)$ の結果に対する利益分配

name	destination	type
BayTours $(s'_{11} + s''_{11}, l'_{11} + l''_{11})$ 30	San Francisco $(s'_{12} + s''_{12}, l'_{12} + l''_{12})$ 20	cable car $(s'_{13} + s''_{13}, l'_{13} + l''_{13})$ 20
BayTours (s'_{21}, l'_{21}) 10	Santa Cruz (s'_{22}, l'_{22}) 10	bus (s'_{23}, l'_{23}) 10
BayTours (s'_{31}, l'_{31}) 0	Santa Cruz (s'_{32}, l'_{32}) 0	boat (s'_{33}, l'_{33}) 0
BayTours (s'_{41}, l'_{41}) 0	Monterey (s'_{42}, l'_{42}) 0	boat (s'_{43}, l'_{43}) 0
HarborCruz (s''_{21}, l''_{21}) 0	Santa Cruz (s''_{22}, l''_{22}) 0	bus (s''_{23}, l''_{23}) 0

図 23 $(ExternalTours) \cup (OriExternalTours)$ の結果に対する利益分配

$$add(l'_{12}, 4)$$

とできる．
また，二行一列目 “Santa Cruz” には価値 60 が与えられている．この時同様に，

$$add(s'_{22}, \frac{36}{2}) = add(s'_{22}, 18)$$

$$add(s'_{32}, \frac{36}{2}) = add(s'_{32}, 18)$$

$$add(l'_{22}, \frac{24}{2}) = add(l'_{22}, 12)$$

$$add(l'_{32}, \frac{24}{2}) = add(l'_{32}, 12)$$

とできる．
次に図 23 において，一行三列目 “cable car” には価値 20 が与えられている．この時同様に，

$$add(s'_{13}, \frac{12}{2}) = add(s'_{13}, 6)$$

$$add(s''_{13}, \frac{12}{2}) = add(s''_{13}, 6)$$

$$add(l'_{13}, \frac{8}{2}) = add(l'_{13}, 4)$$

$$add(l''_{13}, \frac{8}{2}) = add(l''_{13}, 4)$$

とできる．
また，二行三列目 “bus” には価値 10 が与えられている．この時同様に，

$$add(s'_{23}, 6)$$

$$add(l'_{23}, 4)$$

とできる．

5.2 自然結合

自然結合における価値づけは，結合する片方の一つのセルにもう片方の複数のセルが結合する場合がある．この場合片方の一つのセルの直接来歴注釈，間接来歴注釈が結果の関係に多く現れ，価値が与えられた場合にそのセルの直接来歴注釈，間接来歴注釈に必要以上に価値が分配されてしまう．これを防ぐため，価値の付加は一回のみにする．具体的には，問合せを行った者が実際に利用した全セルの直接来歴注釈，間接来歴注釈成分に含まれる最小単位の直接来歴注釈，間接来歴注釈の集合 (集合は要素の重複は削除する) を S, L とし，その最小単位の直接来歴注釈，間接来歴注釈に均等に価値を付加するようにする．実際に利用したセルがわからない場合は結果の関係の全セルが対象となる．

あらかじめ直接来歴注釈と間接来歴注釈に配分する報酬の比を決めておく．今回は (直接来歴注釈):(間接来歴注釈)= 6 : 4 とする．

また，関数 $count(A)$ は，集合 A 内の要素の個数を返すとする．このとき，価値 W が与えられた結果に対して，

任意の $s(\in S)$ 及び $l(\in L)$ に，

$$add(s, W \cdot 0.6 \cdot \frac{1}{count(S)})$$

$$add(l, W \cdot 0.4 \cdot \frac{1}{count(L)})$$

のように価値を付加する．

具体例として，図 14 に価値 100 を与えたとする．このとき，問合せを行った者が実際に利用したセルが，結果の関係の一行一列目，一行二列目，二行一列目の三つであったとする．この三つのセルから S, L の要素を考えると，

$$S = \{s_{11}, s'_{11}, s_{12}, s'_{21}\}$$

$$L = \{l_{11}, l_{12}, l'_{11}, l'_{21}\}$$

となる．これより，以下のように価値を付加する．

$$add(s_{11}, 100 \cdot 0.6 \cdot \frac{1}{4}) = add(s_{11}, 15)$$

s'_{11}, s_{12}, s'_{21} についても同様.

$$add(l_{11}, 100 \cdot 0.4 \cdot \frac{1}{4}) = add(l_{11}, 10)$$

l_{12}, l'_{11}, l'_{21} についても同様.

5.3 選　　択

選択における価値づけは，得られたデータの各セルが必要なものである場合に価値を与えるようにする．ここで選択において必要でないセルとは，例えば条件式 “(属性) = (値)″ となる条件を満たすセルである．なぜなら問合せを行った者がその条件を指定しているためそのデータが得られることは予めわかっているからである．このことを判断するために条件式を最小の条件式ごとに見ていき，必要なデータを得る条件式であるかどうか，つまりその条件式が得られたデータに価値を与える条件式かどうかを決定し，それに応じてセルごとに価値を考える．

表 2　演算子 $\wedge, \vee$ による○×変化

条件式＼元の AB	××	×○	○×	○○
$A \wedge B$	××	×○	○×	○○
$A \vee B$	○○	○○	○○	○○

表 3　演算子 $\neg$ による○×変化

条件式＼元の X	×	○	X : (属性) $\neq$ (値) の場合
$\neg X$	○	○	×

まず，ある最小の条件式について，“(属性) = (属性)″ の等式と，等式以外の条件式を “データに価値を与える条件式” とする．それ以外の条件式 (“(属性) = (値)″ の等式) を “データに価値を与えない条件式” とする．以下では “データに価値を与える条件式” を○，“データに価値を与えない条件式” を×と表現する．

次に条件式 A, B および X について，それらがもともと○であるか×であるかによって三つの演算子 $\wedge$ (AND), $\vee$ (OR) および $\neg$(NOT) による条件式 A, B および X から得られるデータのうち，条件式 A, B および X それぞれに含まれる属性に属するセルが○か×かを以下のように定める． これに基づいて条件式に含まれる属性に属するセルの価値の有無を決定していく．手順としては選択として与えられた条件式全体を間接来歴注釈計算と同様の手法で分解し，その左式，右式が上の表のいずれかによってその条件式が○か×かを決定する．左式，右式が最小の条件式でない場合は操作を再帰的に繰り返し，その後表 2,3 を元に最小の条件式それぞれが○か×かを決定する．

条件式に対して○，×の決定が終了した後，×となっている条件式の属性に含まれない組のセルには価値 w_i を与える．ただしそのセルが，○となっている条件式に含まれる属性に属するセルの場合には，その条件を満たしている場合のみ価値を与える．

価値 w_i がセル c_i に与えられたとする．選択における間接来歴注釈計算の結果算出されたもの (l_{sub} と表現している) に含まれる間接来歴注釈の集合を L とし，セル c_i の直接来歴注釈，間接来歴注釈成分を $(s_1, l_1 \cdot l_{sub})$ とする．この時，

s_1 に価値 ω_s を，l_1 と L に含まれる間接来歴注釈成分に価値 ω_l を与える．

［例 5］ 図 11 に価値 100 を与えたとする．先ほどと同様に

$$\omega_s : \omega_l = 6 : 4$$

が成り立つとする．

まずはじめに図 11 の条件式を変換した式 $(\neg(destination =' SantaCruz') \vee \neg(type =' boat'))$ の○×を上記の手順によって決定すると，

最小の条件式 $\neg(destination =' SantaCruz')$ 及び $\neg(type ='$

name	destination	type
BayTours $(s'_{11}, l'_{11} \cdot (l'_{12} + l'_{13}))$ 20	San Francisco $(s'_{12}, l'_{12} \cdot (l'_{12} + l'_{13}))$ 20	cable car $(s'_{13}, l'_{13} \cdot (l'_{12} + l'_{13}))$ 10
BayTours $(s'_{21}, l'_{21} \cdot l'_{23})$ 20	Santa Cruz $(s'_{22}, l'_{22} \cdot l'_{23})$ 0	bus (s'_{23}, l'_{23}) 0
BayTours $(s'_{41}, l'_{41} \cdot l'_{42})$ 20	Monterey (s'_{42}, l'_{42}) 10	boat $(s'_{43}, l'_{42} \cdot l'_{43})$ 0

図 24　$\sigma_c(ExternalTours)$ の結果に対する利益分配

$boat'$) は図 3 より○となる．さらにこの二つの条件式は ∨ で二項演算されるので図 2 により両方の最小の条件式が○で決定される．

次に「×となっている条件式の属性に含まれないセル」で，さらにそのセルの中で「○となっている条件式に含まれる属性に属するセルの場合には，その条件を満たしているセル」に価値を与える (二つの条件を満たしていても価値を与えなくても良い)．図 24 において，一行一列 “BayTours” は「×となっている条件式の属性に含まれないセル」となっており，さらに「○となっている条件式に含まれる属性に属するセル」ではないので，価値を与えることができる．ここでこのセルに価値 20 を与える．この時，

$$add(s'_{11}, 12)$$

とできる．間接来歴注釈に関しては，

$$add(l'_{11}, v_1), add(l'_{12}, v_2), add(l'_{13}, v_3)$$

$$v_1 + v_2 + v_3 = 8$$

となるように v_1, v_2, v_3 を決定する．

6. 考　　察

4. で定義した間接来歴注釈について，実際に行われる問合せではかなり冗長になる可能性が高い．そのためもっと容易に間接来歴注釈の計算ができる，表示が大きくなりすぎない方法を考える必要がある．

また，4. ではデータベース問合せ演算のうち，選択，射影，和，自然結合の四つについてのみ定義したが，より現実的な問合せ演算である集合関数についても定義するべきである．そこで本章では特に，平均，最大，最小，総数の四つの演算について簡単に考え方を記しておく．

6.1　平均，総数

平均と総数の二つの演算については，結果として得られたその値を出すのに関わった全ての値が，その値を出すために必要だったと考える．そのため直接来歴注釈については結果として得られたその値を出すのに関わった全ての値の直接来歴注釈の積で表される．間接来歴注釈についても，同様に結果として得られたその値を出すのに関わった全ての値の間接来歴注釈の積で表される．

6.2　最大，最小

最大と最小の二つの演算については，結果として得られたその値の元となった全セルの直接来歴注釈および間接来歴注釈の和が，それぞれ結果の値の直接来歴注釈および間接来歴注釈として表される．その値が最大または最小となるためには，比べた他の値がその値よりも小さかった，または大きかった必要がある，という考えもあるが，仮に最大，最小以外の値が存在しなかったとしても最大，最小の値は変化しないため，それらの値の間接来歴注釈は無視することとする．

このようにして簡単に集合関数について考えることもできる．

また，5. で定義した分配方法は，四つの演算のどれか一つを一回だけ使用した問合せについてしか定義できていないため，非実用的である．そのため任意の問合せ式に対して価値の分配が可能であるような方法を考えるか，問合せ結果の直接来歴注釈，間接来歴注釈のみを利用して価値を分配する方法を考える必要がある．ただし問合せ結果の直接来歴注釈，間接来歴注釈のみを利用して価値を分配する方法については，適切に報酬を分配できないことがある．

また，5. で定義した分配方法は，予め問合せ結果のデータに価値を与えられた場合について考えているため，市場主義に寄っている．どれだけ努力してできたデータであっても利用されなければ価値はないとみなせるので，市場主義によることは本報告書の考えに沿っているが，将来的には原価主義に寄った価値の分配方法も考える必要がある．

7. お わ り に

本報告書では，データの起源および貢献度に基づき，データベース問合せ結果に与えられた利益を分配する方法を考えた．データの起源および貢献度を図る尺度として，2. で紹介した Green ら [5] の，和積の形で結果を算出する “How-Provenance” の考えを，4. でセル単位に拡張して直接来歴注釈および間接来歴注釈を定義し，それらが問合せ演算においてどのように計算されるのかについても定義した．4. 3 節では文献 [3] において挙げられている具体例を，本報告書の方法で実行した場合の例について紹介した．5. では，上記で計算された直接来歴注釈，間接来歴注釈を利用し，四つの演算のいずれかを一回だけ用いた問合せにおける，問合せ結果に与えられた報酬の分配方法を定義した．6. では，現段階での問題点および今後の課題を示した．

8. 謝　　辞

本研究は JSPS 科研費 JP16K12437, JP17H06099 の助成を受けたものです．

文　　献

[1] Peter Buneman, Sanjeev Khanna, and Tan Wang-Chiew. Why and where: A characterization of data provenance. In *International conference on database theory*, pp. 316–330. Springer, 2001.

[2] Ashok K Chandra and Philip M Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proceedings of the ninth annual ACM symposium on Theory of computing*, pp. 77–90. ACM, 1977.

[3] James Cheney, Laura Chiticariu, and Wang-Chiew Tan. Provenance in Databases: Why, How, and Where. *Found. Trends databases*, Vol. 1, No. 4, pp. 379–474, April 2009.

[4] Wolfgang Gatterbauer, Alexandra Meliou, and Dan Suciu.

Default-all is dangerous! *arXiv preprint arXiv:1105.4395*, 2011.

[5] Todd J Green, Grigoris Karvounarakis, and Val Tannen. Provenance semirings. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 31–40. ACM, 2007.

[6] Todd J. Green and Val Tannen. The Semiring Framework for Database Provenance. In *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, PODS '17, pp. 93–99, New York, NY, USA, 2017. ACM.

[7] Grigoris Karvounarakis, Zachary G Ives, and Val Tannen. Querying data provenance. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 951–962. ACM, 2010.

[8] Chao Li, Daniel Yang Li, Gerome Miklau, and Dan Suciu. A Theory of Pricing Private Data. *Commun. ACM*, Vol. 60, No. 12, pp. 79–86, November 2017.

[9] Xijun Li, Jianguo Yao, Xue Liu, and Haibing Guan. A first look at information entropy-based data pricing. In *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*, pp. 2053–2060. IEEE, 2017.

[10] OECD. Exploring the Economics of Personal Data. OECD Digital Economy Papers, Organisation for Economic Co-operation and Development, Paris, April 2013.

[11] Aaron Roth. Technical Perspective: Pricing Information (and Its Implications). *Commun. ACM*, Vol. 60, No. 12, pp. 78–78, November 2017.

[12] Victor Vianu Serge Abiteboul, Richard Hull. *Foundations in Databases*. Addison-Wesley Publishing Company, 1995.

[13] 内閣官房 IT 総合戦略室. AI、IoT 時代におけるデータ活用ワーキンググループ中間とりまとめの概要, 3 月 2017 年.