数理解析研究所講究録
第436巻 1981年 27-46

Vocabulary Building for Database Queries

Yuzuru TANAKA

Faculty of Engineering
Hokkaido University

Sapporo, 060 JAPAN

1. Introduction

The introduction of natural language features into query languages seems to be classified into 3 stages. The first stage is the introduction of the syntactic flexibility of natural languages. In this stage, users have to know much about the definite logical access structure defined by the database system. Queries have to describe a desired logical access path in a procedural manner, using some type of syntax similar to some natural languages. The semantics that users can afford is only a one-to-one correspondence between the names of attributes and the actual attributes of the database. While such a system may be able to optimize queries, it can not work if the instruction about which access path to choose is not given from the user.

The second stage is the introduction of the access flexibility. Such a system can define a variety of virtual access paths as well as an actual access structure. This definition may be given by a semantic network or a logical program such as those with PROLOG. Since they are mostly based on the first order logic, they are not flexible enough to define the semantics of a variety of words including those that modify the meanings of predicates.

In my personal view, the third stage that has not been much studied untill now means the introduction of flexible vocabularies. The most part of the semantics of this kind is able to become completely independent from the actual logical structure of the system.

This paper gives a formal basis for vocabulary semantics and vocabulary building facility. For this goal, chapter 2 develops a formal theory of relations with null values and generalized operations. Based on this, chapter 3 defines a database access space with infinitely many access paths, while chapter 4 shows how the stepwise building of a vocabulary defines the formal semantics of a database.

2. Generalized Operations on Relations

2.1 A partial relation

Let Ω and D be two enumerable sets respectively called an attribute set and a value set, where we assume an empty set ϕ belongs to D while another special value "$\perp$" does not. By $(X \to Y)$ we denote a set of all the functions from a set X to another set Y. A subset of $(\Omega \to D)$ is called a relation over Ω, while a subset of $(\Omega \to D')$ for $D'=D \cup \{\perp\}$ is called a partial relation over Ω. An element of $(\Omega \to D)$ is called a tuple over Ω, while one of $(\Omega \to D')$ is a partial tuple over Ω. An attribute value ϕ means the nonexistence of this attribute, while $\perp$ means that the value is unknown. For a partial relation R over Ω, we define $\omega(R)$ as Ω. A special set $(\phi \to D')$ is considered as a singleton $e=\{\varepsilon\}$, i.e.,

$$e=(\phi \to D')=\{\varepsilon\}.$$

Let f be a function from X to Y, and Z be another set. A restriction of f within Z denoted by $f|_Z$ is a function defined as

$$f|_Z \ \varepsilon\ (Z \to Y),$$

$$\text{and} \quad \begin{array}{ll} A \varepsilon X \cap Z & f|_Z(A)=f(A), \\ A \varepsilon Z - X & f|_Z(A)=\perp. \end{array}$$

For each $R \varepsilon (\Omega \to D')$ and a set X, we define a projection of R onto X as

$$[X]R = \{\mu|_{X \cap \omega(R)} \mid {}^{\forall}\mu \varepsilon R \text{ s.t. } \mu|_{X \cap \omega(R)} \varepsilon (X \cap \omega(R) \to D)\}$$

which is always a relation.

<u>Lemma 2.1</u>

A partial relation R is a relation iff $[\omega(R)]R=R$.

(proof)
Obvious.

Different from the case of relations, it does not always hold for a partial relation R that $[X][Y]R=[X\cap Y]R$. The left hand side imposes the condition that the Y-values should be known, while the right hand side imposes only the certainty of $X\cap Y$-values.

The natural join of two relations r and s is defined as

$$\begin{aligned} r*s = \{\ \mu \mid {}^{\forall}\mu \varepsilon (\omega(r)\cup\omega(s) \to D) \text{ s.t.} \\ \mu|_{\omega(r)} \varepsilon\, r \wedge \mu|_{\omega(s)} \varepsilon\, s \\ \wedge\ \mu|_{\omega(r)\cap\omega(s)} \varepsilon\ (\omega(r)\cap\omega(s) \to D-\{\phi\})\ \}. \end{aligned}$$

This definition imposes the condition that the values of join attributes should be known and existent, i.e., they should not be either ϕ nor $\perp$.

2.2 Grouping of values and generalized projections

Here we formalize the so-called "GROUP BY" operations that are fundamental to enrich our formal semantics of queries.

Let r be a relation over Ω and Ω_h be a set defined as

$$\Omega_h = \Omega \cup \{\ X/Y \mid X,\ Y \subset \Omega\}.$$

An attribute (X/Y) of an extended attribute set Ω_h is read as "X grouped by Y values". An extended value set D_h is defined as

$$D_h = D \cup (\textstyle\bigcup_i 2^{(D_h)^i}).$$

For a subset X of Ω_h and a tuple μ of a relation r over a subset of Ω, we define a partial tuple $\mu^h_{r,X} \varepsilon (X \to D_h \cup \{\perp\})$ as follows;

$${}^{\forall}A \varepsilon\, \Omega \cap X \quad \mu^h_{r,X}(A) = \mu(A),$$

and

$${}^{\forall}V/W \varepsilon X$$

$$\mu^h_{r,X}(V/W) = \begin{cases} \{\nu(V) \mid {}^{\forall}\nu \varepsilon [VW]r \text{ s.t.} \\ \qquad \nu|_W = \mu|_W \text{ and } \nu|_V \varepsilon (V \to D-\{\phi\}) \\ \quad \text{if } \mu|_W \varepsilon (Y \to D-\{\phi\}), \\ \perp \quad \text{otherwise.} \end{cases}$$

For a subset X of Ω_h, flat(X) denotes a subset of Ω defined as

$$\mathrm{flat}(x) = (X \cap \Omega) \cup (\cup_{V/W \varepsilon X}(V \cup W)).$$

A generalized projection of a relation r onto a subset X of Ω_h is defined as

$$[X]_h r = \{ \mu^h_{r,X}|_{\omega(r)_h \cap X} \;\Big|\; {}^\forall \mu \varepsilon r \text{ s.t. } \mu^h_{r,x}|_{\omega(r)_h \cap X} \varepsilon (X \to D_h)\},$$

while a generalized projection of a partial relation R onto a subset X of Ω_h is defined as

$$[X]_h R = [X]_h[\mathrm{flat}(X)]R.$$

Now we consider a set of computable functions F defined as

$$F = \{ g \mid g \text{ is computable}, {}^\exists k \geq 0 \;\; g \varepsilon ((D_h)^k \to D_h \cup \{\perp\}) \}.$$

For an attribute set Ω, let Ω_f be a set defined as

$$\Omega_f = \Omega \cup \{ g(A_1, A_2, \ldots, A_k) \mid {}^\forall i \;\; A_i \varepsilon \Omega, \;\; {}^\forall g \varepsilon F \}.$$

For a subset X of Ω_f and a tuple μ of a relation r over a suset of Ω, we define a partial tuple $\mu^f_{r,X} \varepsilon (X \to D_h \cup \{\perp\})$ as follows;

$${}^\forall A \varepsilon \Omega \cap X \quad \mu^f_{r,X}(A) = \mu(A),$$

and

$${}^\forall g(A_1, A_2, \ldots, A_k) \varepsilon X$$

$$\mu^f_{r,X}(g(A_1, \ldots, A_k)) = \text{if } v \text{ is not undefined then } v \text{ else } \perp,$$

where

$$v = g(\mu^f_{r,X}(A_1), \ldots, \mu^f_{r,X}(A_k)).$$

For a subset X of Ω_f, arg(X) denotes a subset of Ω defined as

$$\mathrm{arg}(x) = (X \cap \Omega) \cup (\cup_{g(Y) \varepsilon X} Y).$$

A generalized projection of a relation r onto a subset X of Ω_f is defined as

$$[X]_f r = \{\mu^f_{r,X}|_{\omega(r)_f \cap X} \;\Big|\; {}^\forall \mu \varepsilon r \text{ s.t. } \mu^f_{r,x}|_{\omega(r)_f \cap X} \varepsilon (X \to D_h) \},$$

while a generalized projection of a partial relation R onto a subset X of Ω_f is defined as

$$[X]_f R = [X]_f [\mathrm{arg}(X)] R.$$

A more general attribute set Ω_g is defined as follows;

(1) $\Omega \subset \Omega_g$,

(2) $^\forall X,\ ^\forall Y \subset \Omega_g \quad X/Y \varepsilon\, \Omega_g$,

(3) $^\forall A_1,\ ^\forall A_2,\ \ldots,\ ^\forall A_k \varepsilon\, \Omega_g \quad ^\forall g \varepsilon\, F$

$g(A_1, A_2, \ldots, A_k) \varepsilon\, \Omega_g$,

(4) only those defined by a finite number of applications of the above rules are elements of Ω_g.

For a partial relation R over a subset of Ω, its projection onto some subset X of Ω_g is defined as

$$[X]_g R = \begin{cases} [X]R & \text{if } X \subset \Omega, \\ [X]_f [\mathrm{arg}(X)]_h [\mathrm{flat}(\mathrm{arg}(X))]_g R & \text{otherwise.} \end{cases}$$

Some important functions are defined below, where S denotes some subset of D.

$$\mathrm{sum}(S) = \begin{cases} \Sigma_{v \varepsilon S}\, v & \text{if S is a set of numbers,} \\ \perp & \text{otherwise.} \end{cases}$$

$$\mathrm{max}(S) = \begin{cases} \mathrm{max}_{v \varepsilon S}\, v & \text{if S has some definite order,} \\ \perp & \text{otherwise.} \end{cases}$$

$$\mathrm{min}(S) = \begin{cases} \mathrm{min}_{v \varepsilon S}\, v & \text{if S has some definite order,} \\ \perp & \text{otherwise.} \end{cases}$$

count(S) = cardinarity of S.

For any attributes A, B, and C in Ω_g, we define average(A ; B/C) that is also an element of Ω_g as

$$\mathrm{average}(A;B/C) = \frac{\mathrm{sum}(\mathrm{sum}(A/BC)/C)}{\mathrm{count}(B/C)} .$$

example 2.1

We show a computation process for [average(A;B/C),C]r of an example relation r over {A, B, C, D}.

r	A	B	C	D
	1	a	c	e
	1	a	c	f
	3	b	c	f
	3	a	c	e
	4	b	c	g
	2	a	d	h
	1	a	d	h
	1	b	d	i

$$
\begin{aligned}
&[\text{average}(A;B/C),C]r\\
&= \left[\frac{\text{sum}(\text{sum}(A/BC)/C)}{\text{count}(B/C)},C\right]_f [\text{sum}(\text{sum}(A/BC)/C),\ \text{count}(B/C),C]_f\\
&\quad [\text{sum}(A/BC)/C,\ B/C,\ C]_h [\text{sum}(A/BC),\ B,\ C]_f [A/BC,\ B,\ C]_h [A,\ B,\ C]r
\end{aligned}
$$

r_1=[A, B, C]r		
1	a	c
3	b	c
3	a	c
4	b	c
2	a	d
1	a	d
1	b	d

r_2=[A/BC, B, C]$_h r_1$		
{1, 3}	a	c
{3, 4}	b	c
{2, 1}	a	d
{1}	b	d

r_3=[sum(A/BC), B, C]$_f r_2$		
4	a	c
7	b	c
3	a	d
1	b	d

r_4=[sum(A/BC)/C, B/C, C]$_h r_3$		
{4, 7}	{a, b}	c
{3, 1}	{a, b}	d

r_5=[sum(sum(A/BC)/C), count(B/C), C]$_f r_4$		
11	2	c
4	2	d

$r_6=\left[\frac{\text{sum(sum(A/BC)/C)}}{\text{count(B/C)}},\ c\right]r_5$	
5.5	c
2	d

2.3 Generalized restrictions of relations

Let $P(\underline{x})$ be an n-place predicate, $\underline{A}=(A_1, A_2, \ldots, A_n)$ be an n-dimensional vector of attributes in Ω, and $\underline{\tilde{A}}$ be the corresponding set $\{A_1, A_2, \ldots, A_n\}$. A predicate $P(\underline{A})(\mu)$ for $\mu \varepsilon (\Omega \rightarrow D)$ is defined as

$$P(\underline{A})(\mu) \quad \text{iff} \quad ({}^{\forall}A \,\varepsilon\, \underline{\tilde{A}} \;\; \mu(A) \neq \text{undefined} \wedge \mu(A) \neq \bot)$$
$$\wedge\, P(\mu(A_1), \mu(A_2), \ldots, \mu(A_n)).$$

$P(\underline{A})$ is called a predicate scheme of Ω.

A restriction of a partial relation R over some subset of Ω by a predicate scheme $P(\underline{A})$ of Ω is a relation defined as

$$[P(\underline{A})]R = \{ \mu \mid {}^{\forall}\mu \,\varepsilon\, R \text{ s.t. } P(\underline{A})(\mu) \}.$$

Lemma 2.2

$${}^{\forall}X \subset \Omega \qquad [x][P(\underline{A})]R = [X][P(\underline{A})][X \cup \underline{\tilde{A}}]R.$$

(proof)
Obvious from the definition.

A generalized restriction of a partial relation R over some subset of Ω by a scheme $P(\underline{A})$ for $\underline{\tilde{A}} \subset \Omega_g$ is defined as

$${}^{\forall}X \subset \Omega_g \qquad [X]_g[P(\underline{A})]_g R = [X]_g[p(\underline{A})][X \cup \underline{\tilde{A}}]_g R.$$

3. Intensional Relations and an Information Space

For a partial relation R over some subset of Ω, we define a function ${}^{int}R$ from Ω_g to a set of relations as

$$ {}^{int}R = \lambda x.\ [x]_g R. $$

This function is called a intension of R over Ω. A restriction of ${}^{int}R$ by a scheme $P(\underline{A})$ is defined as

$$ [P(\underline{A})]^{int}R = \lambda x.\ [x]_g [P(\underline{A})]^{int}R(x \cup \tilde{\underline{A}}). $$

Theorem 3.1

$$ \forall X \subset \Omega_g \qquad ([P(\underline{A})]^{int}R)(X) = [X]_g [P(\underline{A})]_g R. $$

(proof)

$$ \begin{aligned} [P(\underline{A})]^{int}R &= \lambda x.\ [x]_g [P(\underline{A})]^{int}R(x \cup \tilde{\underline{A}}) \\ &= \lambda x.\ [x]_g [P(\underline{A})] [x \cup \tilde{\underline{A}}]_g R \\ &= \lambda x.\ [x]_g [P(\underline{A})]_g R \quad \text{(from the definition of } [P(\underline{A})]_g). \end{aligned} $$

A relation $([P(\underline{A})]^{int}R)(X)$ can be informally interpreted as all the information about X that can be obtained from R and satisfies the condition $P(\underline{A})$. We define a set of all the intensional relations over Ω as

$$ IR_\Omega = (2^{\Omega_g} \to ER_\Omega), $$

where ER_Ω is a set of extensional relations over Ω defined as

$$ ER_\Omega = \cup_{\Omega' \in \Omega_g} 2^{(\Omega' \to D_h)}. $$

Let L be an enumerable set called a set of labels. We define a set Ω^L as follows;

(1) $\Omega \subset \Omega^L$,

(2) $\forall A \in \Omega^L$, $\forall l \in L$ lA is an element of Ω^L,

(3) $\forall X \subset \Omega^L$, $\forall Y \subset \Omega^L$ (X/Y) is an element of Ω^L,

(4) $\forall A_1, \forall A_2, \ldots, \forall A_n \in \Omega^L$, $g \in F$ $g(A_1, A_2, \ldots, A_n) \in \Omega^L$,

(5) only those defined by a finite number of applications of the above rules constitute Ω^L.

An elementary adjective for Ω is a triple $(P^{n,m}(\underline{x};\underline{y}), \underline{A}, \underline{B})$, where $P^{n,m}(\underline{x};\underline{y})$ is a (n+m)-place predicate and $\underline{A}$, $\underline{B}$ are n and m dimensional vectors of attributes in Ω_g.

For an elementary adjective $\theta = (P(\underline{x};\underline{y}), \underline{A}, \underline{B})$, its inverse θ^- is defined as

$$(P^*(\underline{y};\underline{x}), \underline{B}, \underline{A}),$$

$$\text{where } \forall \underline{x}, \forall \underline{y} \quad P(\underline{x};\underline{y}) \text{ iff } P^*(\underline{y};\underline{x}).$$

Let Θ be some set of elementary adjectives. Each $\theta \varepsilon \Theta$ is considered as a label for attributes and, for simplicity, $\theta = (P(\underline{x};\underline{y}), \underline{A}, \underline{B})$ is denoted by $P(\underline{A};\theta\underline{B})$, where $\theta\underline{B} = (\theta B_1, \theta B_2, \ldots, \theta B_n)$. Besides, $\underline{A}$ and $\underline{B}$ of $\theta = P(\underline{A}; \underline{B})$ is denoted by $\underline{A}_\theta$ and $\underline{B}_\theta$.

For a pair of same dimensional vectors $\underline{A}$ and $\underline{B}$ of attributes in Ω^Θ, a renaming operator $\underline{A}/\underline{B}$ renames attribute name B_i of a relation to a new name A_i, i.e.,

$$(\underline{A}/\underline{B})r = \{\mu^* \mid \forall \mu \varepsilon r\},$$

$$\begin{aligned} \text{where} \quad & \mu^* \varepsilon ((\omega(r) - \underline{\tilde{B}}) \cup \underline{\tilde{A}} \rightarrow D_h), \\ & \text{and } \forall A \varepsilon \omega(r) - \underline{\tilde{B}} \quad \mu^*(A) = \mu(A) \\ & \quad \forall i \text{ s.t. } B_i \varepsilon \underline{\tilde{B}} \cap \omega(r) \quad \mu^*(A_i) = \mu(B_i). \end{aligned}$$

For simplicity, $(\underline{A}/\underline{B})$ is denoted by $(\underline{\tilde{A}}/\underline{\tilde{B}})$ if its meaning is clear.

Def. 3.1

Let $\underline{r}_0$ be an intensional relation over Ω, and Θ be a set of elementary adjectives for Ω. An information space for $(\Omega, \Theta, \underline{r}_0)$ is an intensional relation $\underline{r}$ over Ω^Θ satisfying

(1) $\forall X \subset \Omega_g \quad \underline{r}(X) = \underline{r}_0(X)$,

(2) $\forall \theta = P_\theta(\underline{A};\theta\underline{B}) \quad \forall X \text{ s.t. } X \cap \theta(\Omega^\Theta) = \phi$

$$\underline{r}(X\theta(Y)) = [X\theta(Y)]_g [P_\theta(\underline{A};\theta\underline{B})]_g (\underline{r}(X \cup \underline{\tilde{A}}) * \underline{r}(\theta(Y) \cup \theta(\underline{\tilde{B}}))),$$

(3) $\forall \theta \varepsilon \Theta \quad \underline{r}(\theta(Y)) \subset (\theta Y/Y)\underline{r}(Y)$,

(4) for each subset X of Ω^Θ, $\underline{r}(X)$ is the maximal set satisfying the above conditions.

Theorem 3.2

The condition (2)(3)(4) of Def. 3.1 may be replaced by the following condition.

$$ {}^{\forall}\theta = P(\underline{A};\theta\underline{B}) \quad {}^{\forall}X \text{ s.t. } X\wedge\theta(\Omega^{\Theta})=\phi \quad {}^{\forall}Y $$

$$ \underline{r}(X\theta(Y))=[X\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(\underline{r}(X\cup\underline{\tilde{A}})*\theta\underline{r}(Y\vee\underline{\tilde{B}})), $$

where ${}^{\forall}r\varepsilon ER_{\Omega}$ $\quad \theta r = (\theta(\omega(r))/\omega(r))r.$

(proof)

Let $\underline{s}$ be an intensional relation satisfying the new condition, and $\underline{r}$ be an information space defined by Def. 3.1.

We first prove the fact that

$$ {}^{\forall}X \quad \underline{s}(X)\supset\underline{r}(X) $$

by mathematical induction on rank(X), where rank(X) is defined as follows;

$$ {}^{\forall}X\subset\Omega_g \quad \text{rank}(X) = 0, $$

$$ \text{rank}(XY)=\text{rank}(X)+\text{rank}(Y), $$

$$ \text{rank}(\theta)=\text{rank}(\underline{\tilde{A}}_{\theta})+\text{rank}(\underline{\tilde{B}}_{\theta})+1=1 \quad (\because\ \underline{\tilde{A}}_{\theta},\ \underline{\tilde{B}}_{\theta}\subset\Omega_g\ ), $$

$$ \text{rank}(\theta(X))=\text{rank}(X)+\text{rank}(\theta), $$

$$ \text{rank}(X/Y)=\text{rank}(X)+\text{rank}(Y). $$

For each X satisfying rank(X)=0, i.e., $X\subset\Omega_g$, $\underline{s}(X)$ equals to $\underline{r}(X)$, and hence $\underline{S}(X)\supset\underline{r}(X)$ holds for rank(X)=0. Let us assume that $\underline{s}(Y)\supset\underline{r}(Y)$ holds for any Y whose rank is less than k, and let rank(Z) be k. We can assume without loss of generality that Z is $X\theta(Y)$ with $X\wedge\theta(\Omega^{\Theta})=\phi$. Then $s(X\theta(Y))$ is equal to

$$ [X\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(\underline{s}(X\cup\underline{\tilde{A}})*\theta\underline{s}(Y\cup\underline{\tilde{B}})). $$

Since it holds that

$$ k = \text{rank}(X\theta(Y)) = \text{rank}(X)+\text{rank}(Y)+1, $$

$$ \text{rank}(X\cup\underline{\tilde{A}}) = \text{rank}(X)+\text{rank}(\underline{\tilde{A}}) = \text{rank}(X)\lneq k, $$

$$ \text{rank}(Y\cup\underline{\tilde{B}}) = \text{rank}(Y)+\text{rank}(\underline{\tilde{B}}) = \text{rank}(Y)\lneq k, $$

it can be assumed that

$$ \underline{s}(X\cup\underline{\tilde{A}})\supset\underline{r}(X\cup\underline{\tilde{A}}), $$

$$ \underline{s}(Y\cup\tilde{B})\supset\underline{r}(Y\cup\underline{\tilde{B}}). $$

Hence it holds that

$$\underline{s}(X\theta(Y)) \supset [X\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(\underline{r}(X\vee\tilde{\underline{A}})*\theta r(Y\vee\tilde{\underline{B}})).$$

From the condition (4) of Def.3.1, $\theta\underline{r}(Y\vee\tilde{\underline{B}})$ must be a super set of $\underline{r}(\theta(Y)\vee\theta(\tilde{\underline{B}}))$. Hence we can conclude that

$$\begin{aligned}\underline{s}(X\theta(Y)) &\supset [X\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(r(X\vee\tilde{\underline{A}})*\theta r(Y\vee\tilde{\underline{B}}))\\ &\supset [\qquad]_g[\qquad\qquad]_g(r(X\vee\tilde{\underline{A}})*\underline{r}(\theta(Y)\vee\theta(\tilde{\underline{B}})))\\ &=\underline{r}(X\theta(Y)).\end{aligned}$$

Since we have proved that, for any X, $\underline{s}(X)$ includes $\underline{r}(X)$, $\underline{s}$ is an information space if it satisfies the conditions (1)(2)(3) of Def.3.1. Obviously, $\underline{s}$ satisfies the condition (1). Let us check the condition (3). From the definition of $\underline{s}$, $\underline{s}(\theta(Y))$ is equal to

$$[\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(\underline{s}(\tilde{\underline{A}})*\theta s(Y\vee\tilde{\underline{B}})),$$

which is included by

$$\begin{aligned}&[\theta(Y)]_g\theta\underline{s}(Y\vee\tilde{\underline{B}})\\ &\subset \theta s(Y).\end{aligned}$$

This proves the condition (3). Since $\underline{s}(\theta(Y)\vee\theta(\tilde{\underline{B}}))$ is

$$[\theta(Y)\vee\theta(\tilde{\underline{B}})]_g[P(\underline{A};\theta(\tilde{\underline{B}}))]_g(s(\underline{A})*\theta\underline{s}(Y\vee\tilde{\underline{B}}))$$

from the definition of this theorem, the substitution of $\underline{r}(X\vee\tilde{\underline{A}})$ and $r(\theta(Y)\vee\theta(\tilde{\underline{B}}))$ in the condition (2) by $\underline{s}(X\vee\tilde{\underline{A}})$ and the expression given above gives

$$\begin{aligned}&[X\theta(Y)]_g[P(\underline{A};\theta(\tilde{\underline{B}}))]_g(s(X\vee\tilde{\underline{A}})\\ &\qquad\qquad *[\theta(Y)\vee\theta(\tilde{\underline{B}})]_g[P(\underline{A};\theta(\underline{B}))]_g(\underline{s}(\tilde{\underline{A}})*\theta\underline{s}(Y\vee\tilde{\underline{B}})))\\ &= [X\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(\underline{s}(X\vee\tilde{\underline{A}})*\underline{s}(\tilde{\underline{A}})*\theta\underline{s}(Y\vee\tilde{\underline{B}}))\\ &= [X\theta(Y)]_g[P(\underline{A};\theta(\underline{B}))]_g(s(x\vee\tilde{\underline{A}})*\theta\underline{s}(Y\vee\tilde{\underline{B}}))\\ &= \underline{s}(X\theta(Y)).\end{aligned}$$

Hence $\underline{s}$ satisfies Def 3.1, and we have proved the theorem.

Theorem 3.3

An infomation space $\underline{r}$ does not depend on the way to choose θ in the condition (2) of Def.3.1.

(proof)

We prove, for X satisfying $X \cap \sigma(\Omega^{\Theta}) = X \cap \tau(\Omega^{\Theta}) = \phi$, $\underline{r}(X\sigma(Y)\tau(Z))$ is uniquely evaluated independently from which adjective should be chosen first.

If σ is first chosen then

$$\underline{r}(X\sigma(Y)\tau(Z))=[X\sigma(Y)\tau(Z)]_g[P(\underline{A};\sigma(\underline{B}))]_g(\underline{r}(X \vee \underline{\tilde{A}} \vee \tau(Z))*\underline{r}(\sigma(Y)\sigma(\underline{\tilde{B}}))).$$

Since $\underline{\tilde{A}} \subset \Omega_g$ holds, $(X \vee \underline{\tilde{A}})$ and $\tau(\Omega^{\Theta})$ are disjoint. Hence $\underline{r}(X \vee \underline{\tilde{A}} \vee \tau(Z))$ is equal to

$$[X\underline{\tilde{A}}\tau(Z)]_g[Q(\underline{C};\tau(\underline{D}))]_g(r(X \vee \underline{\tilde{A}} \vee \underline{\tilde{C}})*r(\tau(Z) \vee \tau(\underline{\tilde{D}}))).$$

Therefore, $\underline{r}(X\sigma(Y)\tau(Z))$ is

$$[X\sigma(Y)\tau(Z)]_g[P(\underline{A};\sigma(\underline{B}))]_g$$
$$(([X\underline{\tilde{A}}\tau(Z)]_g[Q(C;\tau(\underline{D}))]_g(\underline{r}(X \vee \underline{\tilde{A}} \vee \underline{\tilde{C}})*\underline{r}(\tau(Z) \vee \tau(\underline{\tilde{D}})))*\underline{r}(\sigma(Y)\sigma(\underline{\tilde{B}}))),$$

which is equal to

$$[X\sigma(Y)\tau(Z)]_g[P(\underline{A};\sigma(\underline{B}))]_g[Q(\underline{C};\tau(\underline{D}))]_g$$
$$(\underline{r}(X \vee \underline{\tilde{A}} \vee \underline{\tilde{C}})*\underline{r}(\tau(Z) \vee \tau(\underline{\tilde{D}}))*\underline{r}(\sigma(Y)\sigma(\underline{\tilde{B}})))$$
$$= [X\sigma(Y)\tau(Z)]_g[P(\underline{A};\sigma(\underline{B})) \wedge Q(\underline{C};\tau(\underline{D}))]_g$$
$$(\underline{r}(X \vee \underline{\tilde{A}} \vee \underline{\tilde{C}})*\underline{r}(\tau(Z) \vee \tau(\underline{\tilde{D}}))*\underline{r}(\sigma(Y)\sigma(\underline{\tilde{B}})))$$

from the definition of a generalized projection. The last expression is independent of the order between σ and τ.

Let Ω_θ, $\overline{\Omega}_\theta$ be

$$\Omega_\theta = \theta(\Omega^{\Theta})$$
$$\overline{\Omega}_\theta = \Omega^{\Theta} - \Omega_\theta.$$

For each $\theta \varepsilon \Theta$, θ^* and $\overline{\theta}^*$ are defined as

$$\theta^* = \lambda x.\ y \text{ s.t. } \theta(y)=x \cap \Omega_\theta$$
$$\overline{\theta}^* = \lambda x.\ x \cap \overline{\Omega}_\theta.$$

Besides, for each intensional relation $\underline{r}$ and $\theta \,\varepsilon\, \Theta$, $\underline{r}$ is defined as

$$\theta\underline{r} = \lambda x.\ \theta\underline{r}(x).$$

Theorem 3.4

An information space $\underline{r}$ for $(\Omega,\ \Theta,\ \underline{r}_0)$ satisfies the following;

$${}^{\forall}\theta = P(\underline{A};\theta(\underline{B}))\ {}^{\forall}X \text{ s.t. } X \wedge \theta(\Omega^{\Theta}) \neq \phi$$

$$\underline{r}(X) = ([\theta]\ ((\theta\underline{r}\theta^*) * (\underline{r}\bar{\theta}^*)))(X).$$

(proof)
Obvious.

For an information space $\underline{r}$ for $(\Omega,\ \Theta,\ \underline{r}_0)$, we define Θ^* as follows;

(1) $\Theta \subset \Theta^*$

(2) ${}^{\forall}\theta \,\varepsilon\, \Theta \quad \theta^{-} \,\varepsilon\, \Theta^*$

(3) ${}^{\forall}\sigma, {}^{\forall}\tau \,\varepsilon\, \Theta^*$

$\sigma \circ \tau \,\varepsilon\, \Theta^* \qquad (\sigma\circ\tau)^{-} = \sigma^{-}\circ\tau^{-}$

$\sigma + \tau \,\varepsilon\, \Theta^* \qquad (\sigma+\tau)^{-} = \sigma^{-}+\tau^{-}$

$\sigma - \tau \,\varepsilon\, \Theta^* \qquad (\sigma-\tau)^{-} = \sigma^{-}-\tau^{-}$

$\sigma * \tau \,\varepsilon\, \Theta^* \qquad (\sigma*\tau)^{-} = \sigma^{-}*\tau^{-}$

(4) only those defined by a finite number of applications of the above rules constitute Θ^*.

Besides, we define Ω^* as follows;

(1) $\Omega^{\Theta} \subset \Omega^*$

(2) ${}^{\forall}X,\ {}^{\forall}Y \subset \Omega^* \quad X/Y \,\varepsilon\, \Omega^*$

(3) ${}^{\forall}A_1,\ {}^{\forall}A_2,\ \ldots,\ {}^{\forall}A_n \,\varepsilon\, \Omega^*\ {}^{\forall}g \,\varepsilon\, F \quad g(A_1,\ A_2,\ \ldots,\ A_n) \,\varepsilon\, \Omega^*$

(4) ${}^{\forall}A \,\varepsilon\, \Omega^*\ {}^{\forall}\theta \,\varepsilon\, \Theta \quad \theta A \,\varepsilon\, \Omega^*$

(5) only those defined by a finite number of applications of the above rules constitute Ω^*.

An intensional relation $\underline{r}^*$ over Ω^* is defined as

(1) ${}^{\forall}X \subset \Omega^{\Theta} \quad \underline{r}^*(X) = \underline{r}(X)$

(2) ${}^{\forall}\theta \,\varepsilon\, \Theta\ \ {}^{\forall}X \text{ s.t. } X \wedge \theta(\Omega^*) \neq \phi$

$$\underline{r}^*(X) = ([\theta]((\theta\underline{r}^*\theta^*) * (\underline{r}^*\bar{\theta}^*)))(X)$$

where the domains of all the operations are assumed to be extended from Ω^{Θ} to Ω^*

(3) $\forall(\sigma\circ\tau)\ \varepsilon\ \Theta^*$ $\quad \underline{r}^*(X(\sigma\circ\tau)(Y))$

$= (X(\sigma\circ\tau)(Y)/X\sigma(\tau(Y)))\underline{r}^*(X\sigma(\tau(Y))),$

(4) $\forall(\sigma+\tau)\ \varepsilon\ \Theta^*$ $\quad \underline{r}^*(X(\sigma+\tau)(Y))$

$= (X(\sigma+\tau)(Y)/X\sigma(Y))\underline{r}^*(X\sigma(Y))$

$\cup (X(\sigma+\tau)(Y)/X\tau(Y))\underline{r}^*(X\tau(Y)),$

(5) $\forall(\sigma-\tau)\ \varepsilon\ \Theta^*$ $\quad \underline{r}^*(X(\sigma-\tau)(Y))$

$= (X(\sigma-\tau)(Y)/X\sigma(Y))\underline{r}^*(X\sigma(Y))$

$-(X(\sigma-\tau)(Y)/X\tau(Y))\underline{r}^*(X\tau(Y)),$

(6) $\forall(\sigma*\tau)\ \varepsilon\ \Theta^*$ $\quad \underline{r}^*(X(\sigma*\tau)(Y))$

$= (X(\sigma*\tau)(Y)/X\sigma(Y))\underline{r}^*(X\sigma(Y))$

$\cap (X(\sigma*\tau)(Y)/X\tau(Y))\underline{r}^*(X\tau(Y)).$

4. Stepwise Vocabulary Building for Database Queries.

Def. 4.1

A stepwise vocabulary building process for a database with a partial relation Δ over a finite attribute set Ω_0 is a sequence of triples $(\Omega_i, \Theta_i, \underline{r}_i)$ defined as below;

(1) $\Omega_0 \subset \Omega_1 \subset \cdots \subset \Omega_n = \Omega^\infty,$

$\phi = \Theta_0 \subset \Theta_1 \subset \cdots \subset \Theta_n = \Theta^\infty,$

$\Delta = \underline{r}_0, \underline{r}_1, \cdots, \underline{r}_n = \underline{r}^\infty,$

(2) Θ_i is a set of elementary adjectives of Ω_{i-1},

(3) Ω_i is a set $\Omega_{i-1}{}^*$ for $(\Omega_{i-1}, \Theta_{i-1}{}^*, \underline{r}_{i-1})$,

(4) $\underline{r}_i$ is an information space $\underline{r}^*$ of $(\Omega_{i-1}, \Theta_{i-1}{}^*, \underline{r}_{i-1})$.

Then $\underline{r}^\infty$ is called a universal information space of $(\Omega_0, \{\Theta_i\}, \Delta)$ and $\{\Theta_i\}$ is called the stepwise basic vocabularies of adjectives, and Ω^∞ is called the lexicon.

Theorem 4.1

For each $X \subset \Omega^\infty$, $\underline{r}^\infty(X)$ is computable if X and Δ is finite.

(proof)

This can be proved by mathematical induction on rank(x) that is a modified version of the previous defintion and is defined as follows;

$\forall X \subset \Omega_{0g}$ $\quad$ rank(x)=0

rank(XY)=rank(X)+rank(Y)

$\forall i \ \forall \theta \in \Theta_i$ $\quad$ $\mathrm{rank}(\theta)=\mathrm{rank}(\underline{A}_\theta)+\mathrm{rank}(\underline{B}_\theta)+1$

rank(θ(X))=rank(X)+rank(θ)

rank(X/Y)=rank(X)+rank(Y)+1

rank(g(X))=rank(X)+1

$\mathrm{rank}((\sigma \circ \tau)(X))=\mathrm{rank}(\sigma(\tau(X)))+1$

$\mathrm{rank}((\sigma \bullet \tau)(X))=\mathrm{rank}(\sigma)+\mathrm{rank}(\tau)+1$

where • is a one among +, -, *.

$\mathrm{rank}(\theta^-)=\mathrm{rank}(\theta)+1.$

Here we list up several important forms of adjectives.

Def. 4.2

$$^{\forall}A \varepsilon \Omega^{\infty} \qquad \hat{A} = ((x=y)(x;y),\ A,\ A),\ \text{i.e.},\ (A = \hat{A}A).$$

Def. 4.3

$$^{\forall}A \varepsilon \Omega^{\infty} \quad ^{\forall}v \varepsilon D_h$$

$$A^{(\partial v)} = ((y='v')(x;y),\ \phi,A),\ \text{i.e.},\ (A^{(\partial v)}A\partial'v'),$$

where ∂ is one of the relational operators =, ≠, $\geq$, $\leq$, >, <.

Let O be a special attribute that does not belong to Ω^{∞}. For the convenience, we add O to Ω_0 by extending Δ to a cartesian product of Δ and $U = (\{O\} \rightarrow D_h)$, i.e., $\Delta_{new} = \Delta_{old} \times (\{O\} \rightarrow D_h)$. Now we are able to define another important type of adjectives.

Def. 4.4

$$^{\forall}A \varepsilon \Omega^{\infty} \qquad A^{O} = ((x=y)(x;y),\ A,\ O),\ \text{i.e.},\ (A^{O}O = A).$$

We are also allowed to define,for each pair of attributes A and B in Ω^{∞}, their addition, subtraction, and multiplication as below.

Def. 4.5

$$^{\forall}A,\ ^{\forall}B \varepsilon \Omega^{\infty} \qquad A+B = (A^{O}+B^{O})O$$

$$A-B = (A^{O}-B^{O})O$$

$$A*B = (A^{O}*B^{O})O.$$

Theorem 4.2

$$^{\forall}A \varepsilon \Omega^{\infty} \ ^{\forall}X \subset \Omega^{\infty} \qquad \underline{r}^{\infty}((A^{O}O)X) = \underline{r}^{\infty}(AX)$$

(proof)
Obvious.

A vocabulary for Δ is a suset V of Σ^* and α, where Σ^* is a set of all the finite strings of an alphabet Σ and α is a function from $(V \cup \Omega^{\infty} \cup \Theta^{\infty})$ to $\Omega^{\infty} \cup \Theta^{\infty}$ such that $^{\forall}A \varepsilon \Omega^{\infty} \cup \Theta^{\infty} \quad \alpha(A)=A$.

We define for each subset X of V, a relation over X with respect to $(\Omega, \{\Theta_i\}, \Delta)$ as

$$<X> = (X/\alpha(X))\underline{r}^{\infty}(\alpha(X)).$$

example 4.1

For an example relation with

$$\Omega_0 = \{\text{driver, licence \#, typist, typespeed, salary}\},$$

we are able to define the word "employee" as

$$\alpha(\text{employee}) = \text{driver+typist}.$$

For example, <employee, type speed> can be evaluated as

<employee, type speed>
= (employee, typespeed/(driver+typist), typespeed)
$\underline{r}^{\infty}$(driver+typist, typespeed)

= (employee, type speed/(driver+typist), type speed)
((driver+typist), type speed/(driver^{O}+typist^{O})O, type speed)
$\underline{r}^{\infty}$((driver^{O}+typist^{O})O, type speed)
= (employee, type speed/(driver^{O}+typist^{O})O, type speed)
(((driver^{O}+typist^{O})O,type speed/driver^{O}O,type speed)
$\underline{r}^{\infty}$(driver^{O}O, type speed)
$\cup$((driver^{O}+typist^{O})O,type speed/typist^{O}O,type speed)
$\underline{r}^{\infty}$(typist^{O}O, type speed))
= (employee, type speed/ driver, type speed)
$\underline{r}^{\infty}$(driver, type speed)
$\cup$(employee, type speed/ typist, type speed)
$\underline{r}^{\infty}$(typist, type speed)
= (employee, type speed/typist, type speed)
$\underline{r}^{\infty}$(typist, type speed),

while <employee, salary> can be evaluated as

(employee, salary/ driver, salary)$\underline{r}^{\infty}$(driver, salary)
$\cup$ (employee, salary/ typist, salary)$\underline{r}^{\infty}$(typist,salary).

example 4.2

Let us see another database with

$$\Omega_0 = \{\text{employee, department, salary}\},$$

and define an elementary adjective "rich_man" as

$$\text{rich\_man} = (\text{rich\_man(salary)} > \text{average(salary; employee}/\phi)).$$

Then, for example, the query that requests the listing of all the departments that have at least one employee whose salary is more than the average of the company can be simply expressed as

<rich_man(dept)>,

while the list of all such employees can be requested by

<rich_man(employee)>.

5. Concluding Remarks

The vocabulary building facility concerned in this paper is a new cocept of query semantics. It makes a new approach to the effective enhancement of the database usability. The framework developped in this paper has the following features:

(1) Ad hoc vocabulary building is allowed.

(2) As shown in the examples, even a very complicated request is expressed as a very simple query.

(3) It is not necessary for us to describe either a virtual access path nor an actual one.

(4) It provides generalized projections and generalized restrictions.

(5) If we consider a vocabulary V as an attribute set, we can construct a new vocabulary V* over V. Since the description of V* includes no elements of Ω^{∞}, the definition of V* is independent from the database.

(6) For each pair of different concepts, the framework provides the means to define their common concept, their difference concept, and their union concept.

If a proper part of our common vocabulary used in daily conversation is adequately built into a vocabulary of the system, then our man-machine communication will become much smoother and more reliable. By having a common vocabulary, a man and a machine can communicate even a very complicated command with a very few words. Our approach will open out new vistas of these possibilities.