

知識ベース管理システム Kappa の概要

Overview of the Knowledge Base Management System (Kappa)

河村 元夫*
Moto Kawamura

横田 一正*
Kazumasa Yokota

金枝上 敦史*
Atsushi Kanaegami

* (財)新世代コンピュータ技術開発機構
Institute for New Generation Computer Technology

1989年2月10日 受理

Keywords: knowledge base, nested relational model, deductive database, deductive and object-oriented database.

1. はじめに

FGCS プロジェクトのターゲットとして多くの知識情報処理システム (KIPS) が研究開発されつつある。Kappa (Knowledge APPLication-oriented Advanced Database and Knowledge Base Management System) プロジェクト⁽¹⁾は、これら KIPS にデータベース、知識ベースの実験的な環境を提供することを目的とし、1985年9月に ICOT の KBMS プロジェクトの一つとして始まった。動作環境は、PSI/SIMPOS の逐次推論マシンの環境と、PIM (またはマルチ PSI)/PIMOS の並列推論マシンの環境である。前者の環境下で試作版 Kappa-I の実装を 1987年8月に完了した。それは、単なる研究開発の実験としてではなく、ツールとしてすでに種々の応用に利用されている。また、その改良版 Kappa-II を 1989年中にリリースする予定である。

本稿では、2節でシステムの設計方針、3節で知識ベースのためのデータベース、4節で演繹データベースについて述べ、5節で演繹+オブジェクト指向データベースについて議論する。

2. Kappa システムの設計方針

2.1 代表的応用

Kappa の設計に際しては、我々の環境における多くの KIPS を分析し、その中から FGCS プロジェクトの核システムとして期待される、以下の二つの応用の要求を設計に取り入れることにした。

(1) 証明支援システム

これは、証明検証システム、項書換えシステム、証明コンパイラから構成され、さまざまな問題解決システムの核として期待されている。このシステムは種々の数学知識ベースを必要としている。知識としては、証明記述言語で記述された公理、定義、定理、証明がテキストで、証明検証中の証明木や、公理、定理などから得られた推論規則などが項で表現される。またその単位は理論で、知識ベースは理論間の参照関係によって有向非巡回グラフで表現される。また、証明の行間補間のための前向きと後ろ向きの推論は、演繹データベースの問合せ処理に対応している。

(2) 自然言語処理システム

これは多くの KIPS で重要な役割を果たすが、そのために多種で大量の電子化辞書を必要としている。次の三種類の辞書が現在作られつつある。

単語辞書: 数十万の単語を含んだ何種類かの辞書からなり、それぞれのデータ構造は典型的な非正規関係 (後述) である。

概念辞書: 数十万の概念が IS-A 関係の分類階層として表現される。

シソーラス: 単語辞書と概念辞書の間形態で意味ネットワークとして表現される。

2.2 Kappa システムの設計方針

上記の要求を考慮し、以下の設計方針を設定した。

- ① 知識ベースはデータベースの機能を含むべきで、それをデータベースの拡張として位置づける。
- ② システムを次の三つの層から構成する。データベース層、知識ベース層、ユーザ・インタフェース層。応

用に応じて各層を使い分けることが可能である。

- ③ データベース層：基盤となるデータモデルとして非正規関係モデル⁽²⁾を採用し、意味ネットワークや分類階層などを非正規関係モデル上に作る。項をデータ型として扱い単一化やパタンマッチによる検索機能を提供する。大量データを効率的に処理できなければならない。
- ④ 知識ベース層：演繹+オブジェクト指向データベースの枠組に基づく知識表現言語を設計し、その実験的な実装を行う。
- ⑤ ユーザ・インタフェース層：簡単に利用できるインタフェースの実験的な実装を行う。例えば構造データの操作ツールや半自動設計ツール、構造エディタを利用したメタデータ保守ツールなどである。

3. 知識ベースのためのデータベース

データベース層を2節の設計方針に基づいて設計した。ここではデータモデルとその意味を定義し、次にDBMSとしての特徴を述べる。

3.1 非正規関係

非正規関係の形式的表現と意味を定義しよう。

CRL⁽³⁾の記法を使用し、まずオブジェクトを定義する。空オブジェクト ω を含む原子オブジェクトの集合、属性名の集合を仮定し、 $[,]$ を組構成子、 $\{, \}$ を集合構成子とすると、オブジェクトは次のように定義される。

- ① 空オブジェクト ω 、任意の原子オブジェクトはオブジェクトである。
- ② o_1, \dots, o_n がオブジェクトならば、 $\{o_1, \dots, o_n\}$ はオブジェクト(集合オブジェクト)である。
- ③ a_1, \dots, a_n が異なる属性名で o_1, \dots, o_n がオブジェクトならば、 $[a_1/o_1, \dots, a_n/o_n]$ はオブジェクト(組オブジェクト)である。

次の条件を満たす組オブジェクトが非正規組である。

- ① 含まれる各集合オブジェクトについて、そのすべての集合要素には互換性がある。
- ② 属性名に重複がなく、その構造が有限の木領域を構成する。

そして、相互互換の非正規組の集合が非正規関係であり、その要素である非正規組の型(非正規組の原子オブジェクトをそれが属する領域に置き換えたもの)がスキーマであり、属性名としての関係名とオブジェクトとしての非正規関係から構成される組オブジェクトが非正規関係データベースである。

次に非正規関係の意味を考えよう。問題となるのは

非正規関係の特徴である行ネスト操作と行アンネスト操作に対する意味、つまり集合構成子の意味である。例えば、非正規関係 $\{[a/c_1, b/\{c_2, c_3\}], [a/c_1, b/\{c_3, c_4\}]\}$ に対する行ネスト操作の結果を、集合の合併操作に対応させて $\{[a/c_1, b/\{c_2, c_3, c_4\}]\}$ とするが、グルーピング操作に対応させて $\{[a/c_1, b/\{\{c_2, c_3\}, \{c_3, c_4\}\}]\}$ とするかである。我々は前者の立場をとり、非正規関係の意味を行ネスト操作から独立とした。これは参考文献(2)の言葉では表現指向である。これは、意味論としては基本的に関係に基づきながら表現(および蓄積構造)の効率化をめざしているという点で、関係モデルの自然な拡張であり、多値従属性をもつデータの表現と処理の効率化が可能である。この意味を反映して、関係演算を、行、列のネスト、アンネスト操作を含む拡張関係演算として定義しなおした。上記の非正規関係では属性名に一意性の制限をおいているので、問合せ時に行、列のネストの構造を意識する必要がない。

3.2 非正規関係上に実現されるモデル

非正規関係によって、オブジェクト(ノード)間に存在する意味関係を直接表現することはできない。このような構造データを格納するために次の二つのモデルを非正規関係モデル上に実現した。それらは、IS-A や HAS-A などの一種のリンクでノード間が結ばれる分類階層と、ノードが多種のリンクで結ばれたネットワークが多数存在する意味ネットワークである。ただし、これらに対する操作は、ノードやリンクに対するデータ操作のみに限定した。知識表現に依存するリンクの意味を考慮した(推論も含む)操作は、知識ベース層で実現する。

3.3 DBMS の特徴

Kappa-II では非正規関係モデルの実用的な拡張が行われている。

- ① リスト(list)構成子、バッグ(bag)構成子の追加。
- ② データ型としての項と、それらの単一化、パタンマッチ検索。
- ③ 属性間の代数的制約を生成規則、一貫性規則として実現(制約論理型言語 CAL⁽⁸⁾を使用)。
- ④ 非正規関係処理のために有益な操作の提供。

また当然、資源管理、同時実行制御、ユーザ管理などの一般的なデータベース機能を持ち、ほかに次の特徴がある。

- ① 主記憶データベース機能の提供。

- ② ユーザ定義操作の登録と実行.
- ③ 組織別子による中間関係の実現と効率的処理.
- ④ PSI (個人用) が多数結合された分散環境への対応.

4. 演繹データベース

演繹データベースを一口で言えば、論理を基にした、関係データベースの拡張である。言い換えれば関係データベースの証明論による再編成である⁽⁴⁾。我々はこれを知識ベースの第一歩であると考えている。既に確定節に基づく問合せ処理と CRL⁽³⁾ に基づく問合せ処理を実装した。それは横方向情報伝達 (side-ways information passing) による束縛情報の解析、一般化マジック集合法による質問処理の最適化、セミナイーブ評価による上昇評価から成っており⁽⁵⁾、さらに HCT/R⁽⁶⁾ による最適化などいくつかの拡張を予定している。ここでは非正規関係のための論理型言語 CRL の演繹機能について述べる。

4.1 論理型言語 CRL の形式的表現

CRL は、非正規関係の部分クラスである多値関係 (葉属性にのみ集合を許す非正規関係) を処理するために設計された言語である。項 (CRL 項と呼ぶ) は非正規組の定義と同じく属性/値-記法で表現される。

原子オブジェクトの集合、属性名の集合のほか、変数の集合を仮定し、以下のように CRL 項を定義する。

- ① 任意の原子オブジェクトは CRL 項である。
- ①' 変数は CRL 項である。
- ② o_1, \dots, o_n が原子オブジェクトならば, $\{o_1, \dots, o_n\}$ は CRL 項 (集合項) である。
- ③ a_1, \dots, a_n が異なる属性名で t_1, \dots, t_n が CRL 項ならば, $[a_1/t_1, \dots, a_n/t_n]$ は CRL 項 (組項) である。

このように定義される CRL 項の組項のうち、属性名に重複がなく、その構造が有限の木領域を構成するものだけが適切な CRL 項である。

$t, t_1, \dots, t_n (n \geq 0)$ が適切な CRL 項であるとする、 $[head/t, body/\{t_1, \dots, t_n\}]$ または $[head/t]$ の形の組項が CRL 規則であり、それぞれ ' $t \leftarrow t_1, \dots, t_n$ ', ' $t \leftarrow$ ' と書く。また $[body/\{t_1, \dots, t_n\}]$ の形の組項がゴールであり、' $\leftarrow t_1, \dots, t_n$ ' と書く。

CRL 規則の集合が CRL プログラムであり、CRL プログラムが CRL データベースである。これは確定節の演繹データベースの場合と同様に、内包データベース (IDB) と外延データベース (EDB) に分けることができる。このうち EDB は基礎単位節 ($[head/$

$t]$ の形) の集合となるが、実際には属性 *head* をはずした後、非正規関係データベースに格納する。

4.2 CRL の意味論

非正規組の意味に対応し、CRL のモデルは部分タグ木 (PTT) の集合に対応づけられる。PTT は、属性名列からなる木領域 Δ と、 Δ の葉の集合から原子オブジェクトの集合への部分関数 ρ の対で定義される。 M を PTT の集合、 η を変数割当て、 t を適切な CRL 項、 v をアンネスト操作とすると、CRL 項の解釈は次のように定義される。

$$[M; \eta] \models t \leftrightarrow v (\{t \eta\}) \subset M$$

つまり、 t が構造 $[M; \eta]$ で充足可能なのは、 $\{t \eta\}$ のアンネスト結果が M の部分集合であるとき、かつそのときに限る。これは、非正規関係の集合合併操作に基づく意味に対応している。CRL 項は構文上、集合に変数や項を持たないので最汎単一化子が容易に定義できる。

プログラム P のモデル M は、 P に含まれる属性名から構成される PPT の空でない集合であり、次の条件を満たす。

$$\forall q \leftarrow p_1, \dots, p_n \in P (M \models p_1 \wedge \dots \wedge p_n \supset M \models q).$$

この定義から次の基本性質が得られる。

$$q \leftarrow p_1, \dots, p_n$$

$$\equiv q \leftarrow p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_{i-1}, p_{i+1}, \dots, p_n$$

$$\equiv q_1 \leftarrow p_1, \dots, p_n \wedge \dots \wedge q_k \leftarrow p_1, \dots, p_n,$$

ここで p_{i-1}, \dots, p_{i-1} と q_1, \dots, q_k はそれぞれ p_i と q のアンネスト結果を示す。つまりルールをアンネストすることができる。これにより CRL には Prolog のような宣言の意味と手続きの意味、そして最小不動点も定義できる。よって、上昇評価のための最適化も Prolog と同様のアルゴリズムが適用可能であり、CRL 用に定義された拡張 SLD 導出では EDB をアンネストする必要がない。例えば次のゴールを例にとる。

$$\leftarrow G_1, [\dots, a/S, \dots], G_2,$$

ここで S はただ一つの集合項である。サブゴール $[\dots, a/S, \dots]$ が EDB 項 $[\dots, a/S', \dots]$ と単一化子 θ で単一化可能なとき、 $S \setminus S'$ が空集合ならばそのサブゴールは除去され、非空集合ならば、新しいゴールは次のようになる。

$$\leftarrow (G_1, [\dots, a/(S \setminus S'), \dots], G_2) \theta,$$

5. 演繹+オブジェクト指向データベース

KIPS にとって、実世界のオブジェクトをより直接的に表現し、それに対する操作、推論を行うことが重

要である。現在、データベースの拡張として演繹データベース (DD) とオブジェクト指向データベース (OOD) が注目されている。我々はそれらの利点を融合した演繹+オブジェクト指向データベース (DOOD) が、オブジェクトの直接的表現とそれらの DD の枠組での処理を可能にすると考えている。この節では DOOD に向けて検討している枠組の一つ⁽⁷⁾を示す。

5.1 DOOD へのアプローチ

まず我々の考えているオブジェクトと OOD との関係について述べる。OOD はモデル化能力、インピーダンス・ミスマッチの解消、継承の利用など大きな可能性を持つモデルとして考えられている。しかし OOD は明確な定義がされておらず、多様(過剰)な意味が含まれており、またそのアプローチも非形式的なものが多い。‘オブジェクト指向’という言葉には、複合構造、オブジェクト識別、データ抽象、型階層などの静的側面と、メソッド、メッセージ・パッシング、情報のカプセル化などの動的側面がある。その静的側面にまず着目しており、それは非正規関係や複合オブジェクトを含んだものと考えている。

次に、DD の枠組でオブジェクトを操作するにはオブジェクトを形式的に表現する必要がある。このオブジェクトの形式化はいろいろなクラスのオブジェクトに対し行われている。例えば型継承、複合オブジェクト、オブジェクトなどで、4節で示した非正規関係に対する DD もこの一つである。これらのオブジェクトの形式化は構文面ではほとんど同じであり、意味の付与の面でも互いに共通部分や関連部分がある。

これらから、DOOD に対するアプローチとしていくつか考えられるが、現在次の二つを考えている。一つはさまざまな構造データを扱えるように DD を拡張する方向で、他は、オブジェクトに対する質問処理を制約論理型言語スキーム CLP (X)⁽⁸⁾ に対応させる方向である。後者の概要を次節で示す。

5.2 オブジェクトの質問処理スキーム

概要は次のとおりである。一般的なオブジェクトの表現によってさまざまな構造データを表現する。非正規関係や型階層や部品関係などのさまざまな構造データ

は、各問題領域が解釈する具体例(インスタンス)と考える。そして、領域間の媒体はオブジェクトであり、その質問評価は、各領域の評価系によって実行される。通常 CLP (X) の X は個別の領域に具象化されるが、ここでは複数の問題領域を取り扱うことを考えている。

この実現のための基本戦略は次のように考えられる。

- ① オブジェクトは3節での形式化に型を追加する。
- ② 規則は言語に依存した原子型を持つ原子オブジェクトとし、DD は集合オブジェクトとして扱う。
- ③ オブジェクトの形式的意味を ML のような強い型システムによって与える。
- ④ 非正規関係などの種々の構造データは、それぞれ独立した問題領域に対応しており、それらの質問処理は各領域での制約として表現する。
- ⑤ オブジェクトに対する質問は、各問題領域の制約へと分解され、それぞれの制約評価系によって評価された後、それらの組合せを答えとして返す。

このスキームでは、オブジェクトを各意味領域に分解する過程は、論理型言語の導出と対応している。

これには次のような利点がある。

- ① 複雑な構造を単純な構造の組合せと考える。質問についても同様。
- ② 制約論理型言語の枠組で、規則を含んだオブジェクトの評価機構が明瞭になる。
- ③ 新しい応用に対する拡張性に富む。新たな制約評価系を追加するだけでよい。

さらに、オブジェクトの動的側面についても、この枠組で検討中である。

6. おわりに

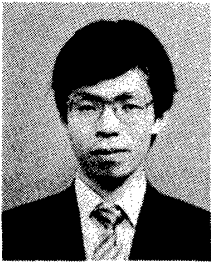
我々は、知識ベースはデータベースの拡張であるべきである、という立場から DOOD の枠組を設定した。また DOOD へのアプローチの一つとして、DD を非正規関係に拡張した。またデータベースの基盤として非正規関係モデルを採用し効率的な処理を実現した。今後、DOOD の研究開発を進めることにより、並列マシン上の Kappa-P、ICOT の関連する他の知識ベースプロジェクト PHI⁽⁹⁾、ETA⁽¹⁰⁾などをこの枠組で統合する予定である。

◇ 参 考 文 献 ◇

- (1) Yokota, K., Kawamura, M. and Kanaegami, A. : Overview of the Knowledge Base Management System (Kappa), FGCS (1988).
- (2) 三浦：非正規データベース理論の動向，情報処理学会アドバンスト・データベース・シンポジウム，(1987. 12)。
- (3) Yokota, K. : Deductive Approach for Nested Rela-

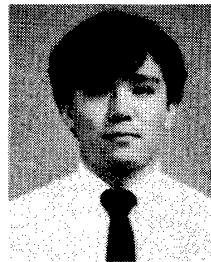
- tions, in *Programming of Future Generation Computers II*, eds. by K. Fuchi and L. Kott, North-Holland (1988).
- (4) Gallaire, H., Minker, J. and Nicolas, L.-M. : Logic and Databases : A Deductive Approach, *ACM Computing Surveys*, Vol. 16, No. 2 (1984).
- (5) 西尾, 楠見 : 演繹データベースにおける再帰的な問い合わせの評価法, *情報処理*, Vol. 29, No. 3 (1988).
- (6) Miyazaki, N., Yokota, K., Haniuda, H. and Itoh, H. : Horn Clause Transformation by Restrictor in Deductive Databases, *ICOT-TR*, 407 (1988).
- (7) 横田 : オブジェクトの形式化とそのスキーム, *電子情報通信学会データ工学研究会*, DE 88-25 (1988. 11).
- (8) 横井, 相場 : 制約ロジック・プログラミング—知識処理への新しいパラダイム—, *情報処理*, Vol. 30, No. 1. (1989).
- (9) 羽生田, 山下, 宮崎, 森田, 伊藤 : 分散知識ベース管理システム PHI, *情報処理学会データベース研究会*, 61-2 (1987. 9).
- (10) Koguchi, T., Kondo, H., Oba, M. and Itoh, H. : Knowledge Representation with Abstractive Layers for Information Retrieval, *FGCS*. (1988).

著 者 紹 介



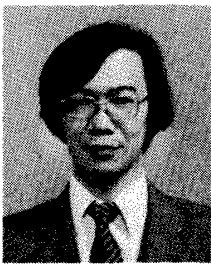
河村 元夫

1984年慶應義塾大学工学部卒業。1986年同大学大学院修士課程修了。同年、日本電気(株)入社。同年、(財)新世代コンピュータ技術開発機構に出向。現在、データベース、知識ベースの研究開発に従事。情報処理学会会員。



金枝上 敦史

1982年大阪大学基礎工学部卒業。1984年大阪大学基礎工学部修士課程物性学修了。同年、三菱電機(株)入社。1985年9月より1989年3月まで(財)新世代コンピュータ技術開発機構に出向。現在、知識ベース管理システムの研究開発に従事。情報処理学会会員。



横田 一正

1972年京都大学理学部(数学)卒業。1985年沖電気工業(株)入社。同年、(財)新世代コンピュータ技術開発機構に出向。現在、データベース、知識ベースの研究開発に従事。データベースと知識ベースへの形式的アプローチ、論理型言語、定理証明などに興味がある。著書に『ゲーデルの世界』(共著、海鳴社)など。情報処理学会、ソフトウェア科学会各会員。
