

TM-0674

知識ベース管理システム

Kappaの概要

河村元夫, 横田 一正,

金枝上敦史

January, 1989

©1989, ICOT

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03) 456-3191~5
Telex ICOT J32964

Institute for New Generation Computer Technology

知識ベース管理システム Kappa の概要

河村 元夫, 横田 一正, 金枝上 敦史

Moto Kawamura, Kazumasa Yokota, Atsushi Kanaegami

(財) 新世代コンピュータ技術開発機構

Institute for New Generation Computer Technology

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、われわれ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

によって非巡回グラフ表現される。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い合わ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 で重要な役割をはたすが、そのために多量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、それぞれは典型的な非正規関係である。

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中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：

- 知識ベースはデータベースの機能を含むべきで、それはデータベースの拡張として位置づける。
- 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。
- 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。

- 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とその実験モジュールから構成される。

- 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と実験的な特徴を持つ。たとえば構造データの操作ツールや半自動設計ツール、構造エディタの利用などである。

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、つぎに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。

CRL³⁾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：

- 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。
- 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、関係名と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。問題となるのは集合構成子で、その意味は、行ネスト操作に対し、集合の合併操作を対応させるか、ベキ集合操作を対応させるかに依存している。われわれは前者の立場をとり、非正規組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。われわれの非正規関係において属性名は一意であるので、問い合わせ時に行、列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

非正規関係はオブジェクト(ノード)間に存在する意味関係を表現することができない。IS-AやHAS-Aなどの一種類のリンクでノード間が結ばれる分類階層と、ノードが多種類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を、非正規関係モデル上に実現した。ただし操作は、ノードやリンクに対するデータ操作に限定した。一般にこれらは、リンクの意味を考慮した(推論も含む)操作が必要だが、それは知識表現に依存するので、知識ベース層で実現される。

3.3 DBMSの特徴

実用のために非正規関係モデルを拡張した：

- リスト構成子、バッグ構成子の追加。
- データ型としての項と、それらの単一化、パターンマッチ検索。
- 代数制約である生成、一貫性規則の実現。
(制約論理型言語 CAL⁴⁾を使用)。
- 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、他に次の特徴がある：
- 主記憶データベース機能の提供。
- ユーザ定義操作の登録と実行。
- 組織別子による中間関係の実現と効率的処理。
- 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構成である⁵⁾。われわれはこれを知識ベースの第一歩であると考えている。既に確立節に基づく問い合わせ処理とCRL³⁾に基づく問い合わせ処理を実装した。それはsideways information passing, generalized magic sets, semi-naive evaluationから成っており⁶⁾、HCT/R⁷⁾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は構文によって対象を非正規関係の部分クラスである多値関係に制限している。項(CRL項と呼ぶ)の定義に非正規関係の定義と同じ属性/値-記法を使う。

原子オブジェクトの集合、属性名の集合の他に、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 ' $t \leftarrow t_1, \dots, t_n$ ' や ' $t \leftarrow$ ' と書く。また $\{t_1, \dots, t_n\}$ の形の集合項がゴールであり、' $\leftarrow q_1, \dots, q_n$ ' 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 を変数割り当て、 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：

$$[M; \eta] \models t \Leftrightarrow v(t\eta) \in M.$$

これは非正規関係の集合合併操作に基づく意味に対応する。CRL項は構文上、集合に変数や項を持たないのでmguが容易に定義できる。単一化は、属性名列の木領域のマージと集合項間の積と原子オブジェクトの識別によって定義される。

プログラム P のモデル M は、 P に含まれるPTTから作られる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：

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

$$\equiv q \leftarrow p_1, \dots, p_{i-1}, p_{i1}, \dots, p_{im}, p_{i+1}, \dots, p_n$$

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

ここで $p_{i1}, \dots, p_{im}$ 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と同様なものが適用可能であり、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。

たとえば次のゴールを例に取る：

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

ここで S は集合項である。サブゴール $[\dots, a/S, \dots]$ が

EDB項 $[\dots, a/S', \dots]$ と単一化子 θ で単一化可能で、かつ $S \setminus S'$ が空集合でないとき、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の要求は、実世界のオブジェクトをより直接的に表現でき、それに対する操作、推論を可能にすることである。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。われわれ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を示す。

5.1 DOODへのアプローチ

まずわれわれ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、オブジェクト同定、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。われわれは、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である。

つぎに、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おこなわれている。たとえば型継承、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次の二つが考えられる。一つは様々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様々な構造データを表現する。そしてオブ

ジェクトに対する処理を、その性質によって限定された問題領域に分解し、それぞれ限定された処理として実行し、それを組み合わせて処理結果を得る。問題領域をCLP(X)の制約評価系に対応させると、この処理はCLP(X)のスキームに対応する。

この実現のための基本戦略は次のように考えられる：

- オブジェクトは3節での形式化に型を追加する。
- ルール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はその集合オブジェクトとして扱う。
- 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。
- 非正規関係などの種々の構造データは意味領域によって具象化される具象物と考える。
- 意味領域はそれに固有の意味と操作を、制約評価系として与えるものである。

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：

- 複雑な構造を単純な構造の組み合わせと考える。質問についても同様。
- 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。
- 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良い。

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

われわれは、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に関連する他の知識ベースプロジェクトPHI、ETAなどをこの枠組で統合する予定である。

参考文献

- 1) K. Yokota, M. Kawamura and A. Kanaegami, "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・シポジウム, 12月, 1987
- 3) K. Yokota, "Deductive Approach for Nested Relations", in *Programming of Future Generation Computers II*, eds. by K. Fuchi and L. Kott, North-Holland, 1988.
- 4) H. Gallaire, J. Minker and L.-M. Nicolas, "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) N. Miyazaki, K. Yokota, H. Haniuda and H. Itoh, "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", 電気情報通信学会データ工学研究会, DE88-25, 11月, 1988
- 8) 横井, 相場, "制約ロジック・プログラミングー知識処理への新しいパラダイム", 情報処理, vol.30, no.1, 1989