

オントロジー工学の試み Towards Ontological Engineering

溝口理一郎

Riichiro Mizoguchi

miz@ei.sanken.osaka-u.ac.jp

大阪大学 産業科学研究所

ISIR, Osaka University

Abstract: We discuss ontology and ontological engineering which is expected to play a role of foundation of so-called Content-Directed AI Research. I first answer questions about what is content-directed research followed by some definitions of an ontology. I next answer further questions about an ontology and describe the merits of ontology design. Finally, I present my idea about Ontological Engineering as well as its future.

1. まえがき

知識処理/AI の分野で大きなパラダイムシフトが起こりつつある。それは次の三つ、(1) 処理中心から情報中心へ、(2) コンピュータ中心から人間中心へ、(3) 形式中心から内容中心へ、に集約される。初めの二つの動向の重要さもさることながら、21世紀に到来する「知識の時代」に最も深く関係する「形式中心から内容中心へ」の傾向は筆者が主張する「内容指向」研究に関するパラダイムシフトである。これまでのAI研究はフォーマリズム指向性が強く、形式的で、一般性が高い研究が中心的な位置を占めている。言い換えれば、「入れ物」を作ることに関わる基礎研究は精力的に行なわれて来たが、何を入れるかという「内容」の議論はその個別性故にないがしろにされてきた。現実の問題を扱うには、対象に関する深い理解とそれに基づく様々な工夫が必要となるが、その議論が欠落しているのである。

筆者は内容指向研究の基礎理論と基盤技術を与える「オントロジー工学」[溝口 97]を提唱すると共に、それを実践してきた。本稿では、オントロジー工学の重要性と意義をわかりやすく解説する[JIPDEC98]。

2. 基本的な問い

2.1 内容指向研究とは何か?

内容指向研究は「コンピュータと人間との知識共有の壮大な試みである」。これまで、形式化可能な「入れ物」の理論と一般性の全くない「領域固有知識」という二分法が広く受け入れられてきた。言い換えると「内容を扱う理論は存在しない」という暗黙の了解があった。内容指向研究はそれらの固定観念を払拭し、21世紀に必要な「知識を積み上げる」こと、そして「知識の相互運用」を可能にする理論と技術を研究する[溝口 96]。

2.2 領域ごとに異なる知識に対してどのような一般理論が可能なのか?

従来学問は領域毎に固有の体系を作り上げてきた。一方、コンピュータの誕生と同時に生まれた情報工学は領域に依存しない「情報」や「データ」という抽象化したものを対象に理論を作り上げてきたが、その成功は領域

の特殊性を捨象した後に残された「もの」を真摯に考察した成果である。内容指向研究は「オントロジー」を核にした適度の抽象化と、対象世界を理解しモデル化する過程をガイドする理論と技術を構築することによって一般理論を構成する(5.1 参照)。

2.3 なぜ知識ベースではだめでオントロジーなのか?

知識は領域依存であり、知識を直接考察の対象にした知識工学は行き詰まり感があることは事実である。しかし、オントロジーは知識を構成する基本概念に立ち戻って考察する。知識の階層性、コンテキスト依存性、知識の分解可能性などを注意深く考察することによって、物事や対象の成り立ちを基本から検討し、知識ベースが持つ様々な問題を回避することの糸口を与える。

2.4 内容指向研究は応用研究にすぎないのでは?

確かに内容指向研究は領域固有の知識を対象にする。しかし、「オントロジー」を武器にして、階層化された適切なサイズの領域内で「知識を積み上げる」理論の確立を目指す。実世界の知識を扱うための基礎理論から技術までを対象とする研究であり、従来の基礎と応用の単純な二分法的考えをうち砕く試みである。

2.5 内容指向研究は形式を軽視するのか?

内容指向研究は「形式指向」偏重傾向を是正する試みであって、決して「形式」技術を排斥しない。むしろその力は物理学における数学のように不可欠である[池田 98]。その顕著な例がオントロジー定義における公理の利用である。内容指向研究はオントロジー工学を核として内容と形式の適度なバランスをとる。

3. オントロジーとは

内容指向研究の核理論・技術としての「オントロジー」は重要な概念であるにも関わらず、それが何を意味するものであるかということに関して多くの混乱が見受けられる。実際、オントロジー関係の会議でもしばしば議論されるが合意が得られてはいないのが現状である。ここでは、いくつかの定義を紹介する。

3.1 いくつかの定義

(1) 本来、オントロジーは哲学の用語であり、「存在に関する体系的な理論(存在論)」という意味を持っており、世の中に存在する全てのものを系統立てて説明することを目指している。なにが存在として本質的で、なにが存在物の区別をするのか? 存在の間の関係は? などと言うことを考え、その結果としての理論と体系がオントロジー(存在論)である。

(2) 人工知能の立場からは、「概念化の明示的な規約」 by Tom Gruber という定義がなされる。ここで概念化とは、対象(世界)に関して、興味を持つ概念とそれらの間の関係とを指す。我々が世界を認識し、そのモデルをコンピュータ内に作るうとするときには、必ず世界を概念化する。例えば、ビジネスプロセスのモデルを構築する際には、ビジネス自体を構成する概念を洗いだし、それらの間の関係を整理し、概念を特徴付け、他と区別する属性を明確にする。そして、それらの概念を記述する。概念のなかには、オブジェクト(名詞に対応)だけではなく、活動を表す動詞的なものもある。そして、活動が対象とするオブジェクトとそれらとを関係付ける制約がある。これら全てを体系化したものがオントロジーである。

(3) 知識ベースの立場からは、「人工システムを構築する際のビルディングブロックとして用いられる基本概念/語彙の体系(理論)」 by Mizoguchi という定義がなされる。知識ベースは対象とする世界における問題解決モデルを構築するとき不可欠となるものである。問題解決を対象とするので、オントロジーは問題解決過程に固有の概念化であるタスクに関するオントロジーであるタスクオントロジー、そしてタスクが実行される領域(ドメイン)に関わるドメインオントロジーの大きく2種類に分かれる。

(4) 緩い定義

「ある目的のための世界の認識に関する共通の合意」 by Takeda (奈良先端大)。ここでは、定義の一般性を増すことに留意して意識的に合意内容を規定することは避けている。もう一つの特徴はオントロジー目的依存性を明示していることである。

(5) もう一つの定義 by Tom Gruber

オントロジーは複数人の間で共有される合意内容である。共有される合意内容としては、ドメイン知識のモデル化の概念フレームワーク、相互に交信し合うエージェントの会話の内容に関わる通信規約、ドメインの知識を記述するときの合意事項、などが含まれる。知識共有の考えのなかでは、オントロジーは表現のための語彙の定義の形で記述される。最も簡単な場合としては、カテゴリーとそれらの間の上下関係(包含関係)を規定したものの(概念階層)であり、関係データベースにおける概念スキーマがある。

3.2 典型的な問い

(1) オントロジーと知識ベースは何が異なるのか?

以下の解答は Ontolingua mailing list において Stanford 大学の Adam Farquhar <axf@HPP.Stanford.EDU> によって、Wed, 26 Feb 1997 に掲載された意見の翻訳である。

(a) それはその分野の人々(エージェント)の合意した知識を表現したものか?

(b) 人々はそれを正確に定義された用語として参照しているか?

(c) その表現に用いられている言語は人々が言いたいことを表現するのに十分な表現力を持っているか?

(d) それは複数の異なった問題解決に再利用可能か?

(e) それは安定しているか?

(f) それは、新しい知識ベース、データベーススキーマ、そしてオブジェクト指向プログラムなどの複数のアプリケーションプログラムを開発するための出発点として利用可能か?

Yes の解答が多ければ(強ければ)多いほど、それだけオントロジーらしくなる。

このように、「オントロジーと知識ベースには明確な境界はない」というのも一つの立場である。それは、オントロジーも言うまでもなく、知識の一種であることから当然の帰結であるとも言える。しかし、「知識ベース」の意味がきわめて曖昧であるのでその定義をすることも必要である。例えば、Cyc のような知識ベースを意味するのであればこの問はまさに的を射た問であるといえる。しかし、専門家の経験則を蓄積したエキスパートシステムのルールベースを想定すれば、オントロジーとの比較は問題外となろう。なぜならそのようなルールベースは開発者が持つ(暗黙的な)オントロジーに基づいて構築されているからであり、決してオントロジーを表現していないからである。

(2) オブジェクト指向方法論、あるいはクラス階層と何が違うのか?

確かにオブジェクト指向言語のクラス階層とオントロジーの概念階層との類似性は高い。また、オブジェクト指向方法論はその上流における分析ではオントロジー開発方法論との共通点は多い。しかし、実装に近いフェーズでは前者は実行性能を重視し、後者は宣言的記述に基づき明示性、形式性を重視する。そして、決定的な相違点は公理の「宣言的」かつ「明示的」記述性の有無にある。すなわち、オブジェクト指向言語は本質的に手続き的であるためクラスの意味、クラス間の関係、そしてメソッドの持つ意味はその手続き的記述の中に埋もれており暗黙的である。一方、オントロジーでは各概念の意味定義と概念間の関係記述は宣言的、かつ明示的である。手続きに関してもその仕様は宣言的に記述され、形式性、明示性は維持される。

(3) 何が新しいのか? Taxonomy と何がちがうのか?

オントロジーは「Taxonomy(概念階層)」という意味を含意するが、それ以外の意味も含んだ概念である。一般に、新しい用語・概念は既存の概念を発展させたものが多く、全く新しい概念であることはまれである。多くはそれまでに存在していた概念を拡張したもの、あるいは新しい観点から再評価したもの、などである場合が多い。オントロジーもその例外ではない。すなわち、従来から存在していた「概念階層」、「標準語彙」、「Upper model」などの概念を含むとともに、知識ベースの開発に際してなされている暗黙の仮定、前提となっている対象世界の概念化(利用目的に依存した概念とそこで興味のある概念間の関係)を述語論理などを用いた明示的かつ形式的な定義という新しい概念が付加されている。

(4) 誰が作るのか?

エキスパートシステムの知識ベースの開発においてもこの種の質問はしばしばなされた。知識ベースを構築するのは知識工学者かドメインエキスパートシステムか、という質問である。どちらでなくてはならないということはない。いずれの場合においても両方の知識が必要であることには変わりはない。

Upper オントロジーは本質的には哲学者の領分であるが、コンピュータ科学者の協力によってコンピュータという「武器」を駆使してこれまでにない、コンピュータ

科学のニーズに合致したオントロジーが構築される可能性がある。

(5) 汎用性の高い上位レベルオントロジーとタスク依存のオントロジーとは両立するか？

工学的にはオントロジーはその利用目的に依存するものであって、汎用のオントロジーは使いものにならないという考えが支配的である。その代表的な考えがタスクオントロジーであり、問題解決に用いられるオントロジーはタスクが要求するように組織化されているべきであるというものである。しかし、一方では Upper モデルと呼ばれる汎用のオントロジーが存在する。両者の距離は遠い。まず思想の相違は一見決定的であるように見える。しかし、両者の用途を考えると、Upper モデルは領域やタスクを越えた多くの人々間で共有することを目指した概念化であり、タスクオントロジーに現れるタスク固有の概念においても仮定される基本的なものである。実際、両者の遠い距離を埋めた例は少ないが、法律オントロジーでの成功に見られるように不可能であるとは思われない[山口 98]。タスクオントロジーは合意可能な Upper モデルを前提としてその上に積み上げる形で設計すべきなのであろう。

(6) オントロジーは単なるラベルの集合か、あるいは実行可能な概念定義も意味するのか？

確かに「オントロジー」が意味するところは広範囲にわたっている。その計算論的意味は極めて幅広い。そこで、筆者はオントロジーに計算論的意味づけの観点から次の3つのレベルを設定することを提案している。

・レベル1

オントロジーの基本的な機能は、対象世界に存在する概念の切り出し（選択）とそれらの関係の記述である。最も一般的で簡単な記述が階層関係の記述であり、そこには概念のラベルと階層記述だけが存在する。このレベルのオントロジーを最もプリミティブなものであるという意味で、レベル1オントロジーと呼ぶ。Internet 上のドキュメントのメタデータ記述に使われるオントロジーや探索エンジンの概念階層はこのオントロジーの典型である。

・レベル2

次に各概念の意味定義（制約）や関係の記述（公理的記述）が加わることによって、オントロジーを利用したモデル構築において種々の適切なガイドや示唆を与えると共に、オントロジーを用いて記述できるもの全体の性質(Competence)に関する質問に答えることができる。このオントロジーをレベル2オントロジーと呼ぼう。多くのオントロジーはこのレベルを目指している。

・レベル3

このレベルのオントロジーはオントロジーを用いて構築されたモデルのある問題解決における実行のパフォーマンス(Performance)、即ち、オントロジーを用いて記述したものが実行されたときの振る舞いに関する質問に回答する。このようなパフォーマンスに関する質問に対しては、本質的に手続き的な記述が必要な場合があり、形式的な公理と証明系では答えることができないことが多いのが現実である。そこで準公理が必要となる。KADS のオントロジー[Breuker94]やタスクオントロジー[テハリノ 93][瀬田 98]はこのレベルのオントロジーである。

(7) 具体的にどのようなメリットがあるのか？

オントロジーの目的依存性から、オントロジーはその利用目的に合致した様々な機能を持っている。ここでは、それらの機能を数え上げる。

(1) 語彙の提供

対象とする世界を記述する際には厳密に定義された、

関係者の合意に基づく「語彙」が必要であるが、オントロジーはそのような標準的な語彙を提供する。

(2) 暗黙情報を明示化

ある領域において仕事をしている人々にとって無意識に仮定しているものが多くある。その典型が日頃頻繁に用いている用語（概念）の定義であり、それらの概念間に存在する基本的な関係や制約である。また、知識ベースは何らかの概念化に基づいているが、その概念化に関する情報は多くの場合暗黙的である。オントロジーはまさにこのような暗黙知識を記述したものであり、それらを明示化する役割を持っている。

ここで注意しなければならないことは、「単に」明示化しただけである、とは思わないことである。この「明示化」がいかに重要であるかは経験した人にしか理解できないと言う宿命的なことがあるが、暗黙の前提を明示化できなかった（しなかった）ことが知識共有・再利用を妨げていた大きな原因の一つであることを認識することは重要である。明示化の効果は計り知れないものがある。

(3) 概念定義

厳密な定義をすることなく日常的に用いている概念の明確な定義を提供する。これによって誤解が減り、人間同士、そしてコンピュータと人間との意味共有の第一歩が始まる。

(4) データ構造

データベースの概念スキーマはデータベースのオントロジーであるが、その意味においてオントロジーは対象とする世界に存在する概念や情報を記述するデータ構造を提供する。

(5) 知識の体系化

知識を体系化するにはまず厳密に定義され、関係者が合意する概念（用語）が必要である。そしてそれらを用いて様々な現象、観測事象、興味ある対象を説明する理論が記述され、組織化される。このようにオントロジーは知識を体系化する際の拠り所となるバックボーンとしての機能を持っている。実際、筆者らは IMS プロジェクトの Gnosis プロジェクトにおいて、設計知識の体系化をタスクオントロジーの概念に基づいて研究を進めている [IMS97][上田 97]。

(6) 標準化

今後の知識処理の高度化、高効率化を考えると、標準化の問題を避けて通ることはできない。工業製品の生産性の高さを考えれば明らかのように、必要な基本部品の標準化は生産性の高度化に貢献するところが大きい。ボルトとナットに対応するような規格品を知識処理の世界でも作る必要がある。オントロジーは標準概念の意味を規定するものであり、オントロジーの一般性、共通性、形式性は標準化に貢献する。

(7) 設計意図

オントロジーはシステム設計者が持っている対象に関する理解、システム設計の意図、すなわち design rationale を明らかにする。オントロジーは前提とされている条件や環境、解くべき問題が要求する仮定などの暗黙的な情報、そしてそれを反映した対象の世界の概念化に関わる根本的な情報を明示し、知識ベースの構築を支えるバックボーンとして機能する。具体的には、診断システムであれば診断可能な故障原因のクラス、推論可能な範囲、定性推論であれば導出可能な因果関係のクラス、ドメインのモデルであれば想定したタスク、など設計者の意図が明らかになる。

(8) メタモデル的機能

モデルは現実の対象を抽象化してコンピュータ内部に作られる。そして、オントロジーはある対象をモデル化

するとき必要となる概念とそれらの間に成立する関係を明示的に規定し、そのモデルはオントロジーが提供する概念と制約の下で作られる。この意味で、オントロジーはメタモデルとしての機能を持っていることができる。

(9) Theory of content (内容の理論)

以上のことを総合して、オントロジーは、とすれば個別の状況に依存したアドホックな議論になりがちで形式理論の様に積み重ねが効くような方法論を持たなかった「内容指向研究」に対して、内容を扱うために必要な新しい「理論」、Theory of Content を提供する(5.1 参照)。

次に、オントロジーの実用的な効果を見るためにオントロジーに基づいて開発されたシステムが持つと期待される利点を考察する。

(1) 再利用性

システム、各モジュールに基づいている世界モデルの明示的な記述はシステムの仮定や前提を明らかにし、その再利用性を向上させる。しかし、再利用にはモジュールの検索、検証、修正などの操作が必要であり、システムの透明性だけでは不十分である。

(2) 相互運用性

上述の前提知識の明示化に加えて、システムの機能や通信プロトコルの記述などが容易になるとともに、オントロジーの形式性が検証を容易にするので相互運用性が向上する。

(3) 実装容易性

対象世界の概念化とその形式的記述はデータの論理構造や概念(オブジェクト)の機能の形式的仕様記述としての役割を果たすため、概念設計結果が詳細かつ厳密に記述され、実装が容易となる。

(4) 知的ツール

ツールはそれが構築するシステムが持つべきモデルに関するモデル(メタモデル)を持っていないと必要ならぬが、オントロジーが持つメタモデルとしての性質はまさしくそのためにあるようなものである。言い換えれば、オントロジーは何を生成するためのツールであるかということを知っている知的なツールの構築を可能にする[Ikeda97]。

4. 研究の状況

4.1 理論

オントロジー理論は N.Guarino と J.Sowa 等によって積極的に進められている。Guarino は FOIS98(Formal Ontology in Information Systems)と呼ばれる、KR98の直後に開催された世界で最初のオントロジーに関する国際会議の議長であり、Sowa は Conceptual Graph[Sowa98]の提案者でありオントロジー理論と並行して知識表現の中間言語標準化を積極的に進めている。

両者に共通する思想は、工学的な応用を常に意識しつつ、哲学の研究成果としての Top-level オントロジーを重視する点である。多くの実践派の人々は Top-level オントロジーの有用性を疑問視する。所詮オントロジーは特定の目的に適合するように設計されていなければ有用性の点で問題が起こり得るからである。しかし、大規模な知識ベースをある特定のドメインで構築することを最終目標としてオントロジーを開発する場合には、その妥当性を特定タスクを越えたある程度の一般性があるところで求めなければならなくなり、その際によりどころとなるのが Top-level オントロジーであろう。哲学分野の成果は考察に十分値すると筆者は考えている。

Soawa の Top-level オントロジーを図1に示す。その

本質は Peirce が提唱した、Firstness (それ自身で定義可能な概念: 人間, 鉄など), Secondness (他の存在物を参照しなければ定義できない概念: 妻, 教師など), Thirdness (Secondness を持つ概念がそれ自身であり得る環境: 家庭, 学校など) に加えて, abstract と physical, そして Continuant (実体が持つべき性質) と Occurrent (プロセスが持つべき性質) の二つを追加して, 合計 7 つの基本 Property を本に 12 個の Top-level カテゴリーを求めている。

図1. Sowa のオントロジー[Sowa95]

Guarino のオントロジーは更に本格的である[Guarino 97]。全体・部分の理論(部分とは何か, どのような性質を満たせば部分であることを認定できるのか, 何種類の部分があるのか, 全体とは何か, どのようなときに全体性は失われるのか, などに答える理論), identity の理論(実体は identity を維持したままどのような変化が許されるのか, どのような時 identity は失われるのかなどの問いに答える理論), 依存関係の理論(依存するとは何か, どのような種類の依存関係があるのかなどに答える理論: 上述の Firstness, secondness などの考えに深く関係する)の3つの哲学的理論の成果を踏まえつつ, 厳密な Top-level オントロジーを設計している。

もう一つの顕著な特徴は, オントロジーを我々人間が認知する実体を表す Particulars (物体, 出来事, 物質など) とそれらについて語るときに必要な概念 Universals (概念, 属性, 状態, 関係など) に2分している点である。そして各々についてオントロジーを設計している。Universals のオントロジーの例を図2に示す。ここで中心となっている概念認識の原理は, identity が matter (物質) のように外延的(部分の単なる集合というだけで identity が認定できる)か person の様に内包的であるか, 永続性か一過性のものか(Rigidity)などの根本的屬性の有無である。

図2 Guarino の Universals のオントロジー [Guarino 97]

図3は Particulars のオントロジーの一部である。大きな特徴は, とすると自然言語に現れる語彙の範囲で抽象度の高い概念を基準にオントロジーを作っていく傾向がある中で, 上述の3つの理論に則って明確な根拠の元に概念選択を行っていること, 特に依存関係理論の成

果の利用が顕著である。たとえば、object は究極的に matter, space, そして time に依存するわけでそれらを substrate としてまとめている点はおもしろい。

なりつつある。MCF は Netscape 社が W3 コンソーシアムに提案したメタデータ自身のデータモデル、構文、語彙などの標準化案である。構文は XML で書かれている。RDF は W3 コンソーシアムで標準化されつつあるメタデータ記述のフレームワークである。これらは筆者のいう、インデキシング(レベル1)オントロジーである。

XML タグセットは特定領域の人々の間で共有することを前提として、対象ドキュメントの意味に踏み込んだ記述を行うのが普通でありインデキシングオントロジーの域を越えてくる可能性がある。たとえば知的教育システムでは、学習者の理解状態に応じて単元の進み方、問題の提示、説明の提示などを適応的に変化させる必要があるが、そのようなシステムの振る舞いを制御することに関わる情報と制御対象を識別するタグなどを開発・共有することによって、XML タグ付け教材を実行するエンジンの仕様を標準化することができる。そのようなエンジンは「Plug & play の概念に基づく知的な Player」と呼ぶことができるが、今後は教育に限らず多くの分野で XML タグとタスクオントロジー、そして Java で実現された知的 Player との関わりは深まるものと思われる。

図3 Guarino の Particulars のオントロジー
[Guarino 97]

4.2 機械可読辞書(MRD)

自然言語処理には大規模な機械可読辞書が必要となるため、自然言語処理の分野では古くから辞書の開発が活発に、情報処理の分野の中では早くからオントロジーの議論が行われていた。前述の定義から辞書とオントロジーが同一であるとは言えないが、最近の MRD はその上位概念の構造を Upper model と呼んでオントロジーと同列に議論ができる様な構成をとっている例が多い。その代表的なものが、WordNet [Miller93] [WordNet], EDR [横井97], EuroWordNet [EWN], Generalized Upper Model [GUM], などであろう。辞書ではないが自然言語処理を意識して構築された常識ベース Cyc [Cyc] も忘れてはならない。ここでは詳しい議論はできないので、詳細は JIPDEC の「オントロジー工学委員会」調査報告書 [JIPDEC98] を参照されたい。

ここで言及しておきたいことは、ANSI ad hoc 委員会にオントロジーの標準化を行う委員会ができており、すでに2年ほど活動が続けられていることである (URL: <http://ksl-web.stanford.edu/onto-std/>)。実現するかどうかは別として、Cyc, WordNet, EDR などの知識ベースや機械可読辞書の Upper model を取り出してそれらの共通性、相違点を洗い出して概念間の対応をとりながら Reference Ontology (RO) と呼べるようなものを作るべく努力を重ねている [Ogino et al. 97]。今のところ、相違点の方が共通性を圧倒しており RO が無事構築される可能性は小さいが、とにかく彼らは果敢にもオントロジーの標準化という困難な問題に取り組んでいることは注目に値する。

4.3 Metadata, XML タグ, そして知的 Player

標準化に関わるもう一つの動きは Dublin Core [DC97], MCF: Meta Content Framework [MCF97] や W3 コンソーシアムの RDF: Resource Description Framework [RDF97] に代表される Internet 上の情報を対象にした Metadata 記述に関する標準化活動である。Dublin Core はそもそもメタデータとしてどのような情報を記述すべきかという基本的なことに関するワークショップが開催されそのときの成果をまとめた物であるが、一種の de fact 標準に

4.4 方法論

オントロジーの開発は容易ではなく、優れた方法論と開発環境が望まれる。ここではいくつかの方法論を紹介する。

4.4.1 IDEF5 [IDEF5]

企業活動やシステムのモデリングと分析技術とツールの標準化を主に行っている KBSI (Knowledge Based Systems, Inc.) は企業活動モデリングを意識したオントロジー開発方法論とオントロジー表現言語、IDEF5 を開発している。オントロジー開発の基本は、語彙の開発にあることを明確にし、用語の抽出、その厳密な意味定義、そして用語間の意味制約を形式的に記述することを方法論の骨子としている。かなりしっかりしたドキュメントもあり、そこには開発の際の注意事項などもかかれているが、オントロジーの図的表現言語を整備しているところが一つの特徴であろう。

4.4.2 TOVE [TOVE]

カナダのトロント大学で開発された TOVE 方法論は IDEF5 と同じく、企業活動モデリングのためのオントロジー開発方法論であるが、ユニークな位置を占めている。基本思想を以下に示す。

- (1) オントロジーに期待する能力を、オントロジーを用いて構築したモデルへの質問 (Competency questions) という形で整理する。すなわち、オントロジーに基づいて構築したモデルに答えてほしい質問を要求仕様の形で整理する。TOVE ではオントロジーはモデルを規定する公理系として機能することが前提となっている。
- (2) オントロジーの公理は Prolog を用いて記述され、上述の質問に答えることによってオントロジーを検証する。

このように要求仕様を明確にしてオントロジーを設計し、それを評価に用いるというソフトウェア開発の正統的な方法論である。企業活動オントロジー開発のために彼らが用意した質問の例は：

- (1) あるゴールを達成するためにはどのような活動が遂行されなければならないか？
- (2) 未来の複数時点における活動が与えられたとき、それ以外の時点における資源と活動に関する性質を聞く。

(3) あるタスクを前に(後ろに)移動するとどのようなことが起こるか?

などである。これからもわかるように TOVE では時間と活動の厳密な定義を公理化して、概念レベルにおけるモデルの実行を通して定性的な範囲でのモデルの振る舞いに関する様々な性質を調べることができる。

4.4.3 DODDLE, アスペクト理論, CLEPE & AFM

我が国におけるオントロジー開発環境を紹介する。DODDLE(A Domain Ontology rapiD DeveLopment Environment) [Yamaguchi97] は静岡大学の山口らによって開発された環境で法律オントロジーの開発に使われた実績がある。一般的な言語知識である WordNet の概念体系を利用して、法律という領域固有オントロジーを開発する際に適切にガイドする機能があることが特徴である。開発されるオントロジーは現状ではレベル1である。

奈良先端大の武田らのアスペクト理論[Takeda95]は同一の世界を複数の視点から見た場合にできる複数のオントロジーを統合する枠組みであり、かつオントロジー構成の方法論を提供している。一つのオントロジーをアスペクトと呼び、それらの間の論理演算を定義して、オントロジーの統合操作を支援する。ASPECTROL という言語が用意されており、エージェントコミュニケーションにおけるオントロジー記述にも利用されている。対象となるオントロジーはレベル1である。

CLEPE は大阪大学の池田らによって開発されている、レベル3のオントロジーであるタスクオントロジーの記述言語・利用環境である[瀬田 98]。3種類のユーザ、すなわち、基本オントロジー開発者、タスクオントロジー開発者、そしてタスクオントロジーを用いてモデルを開発する利用者を想定して言語設計が行われている。モデルはタスクオントロジーの公理の制約の下に作られ、モデルの整合性が維持される。また、モデルは概念レベルでの実行が可能であり TOVE と同様様々な質問に答えることができる。

AFM(Activity-First Method)[Mizoguchi95][久保 98]はタスク依存のドメインオントロジー開発の方法論である。ドキュメントからの開発を想定しており、はじめにタスク分析を行って「動詞概念」を抽出し、動詞が目的語にする「名詞」概念をそのタスクにおける「役割」の下に組織化するという手順を踏む。システムはソースドキュメントから数種類の間生成物を経て最終出力のタスクオントロジーまでの変化を管理しており、きめ細かい支援が可能となっている。レベル2のオントロジー開発を対象にしている。

5. オントロジー工学

本節ではオントロジー工学そのものに関する具体的な問いに答えてみたい。

5.1 「オントロジーは内容の理論である」とはどのような意味か?

物理学が理論科学として見事に成立していることの原因を考えてみると数学の働きが大きいことがわかる。数学には、

- (1) 自然数, 実数, 連続関数, 解析関数
などの厳密に定義可能な変数のクラスと、それを操作したり、関係を記述する
- (2) 四則演算, 微分, 群, 環, 線形性, 直交性
などの演算クラスや関係が定義されている。そして、物理学が対象とする現象を説明するために必要な変量と演算のほとんど全てが数学で既に準備されている概念に

置き換えることができる。このことが物理学を理論として成立させている根本にある。すなわち興味ある現象という「内容」を扱う理論体系が存在するということである。

それでは知識処理の場合はどうであろうか? 理論の典型である述語論理を例に取る。それを知識表現と見なした場合、本来豊富であるべき知識の「内容」を、真理値をとる「述語」という「一つの」カテゴリだけに押し込めて、その見返りとして美しい形式理論を手に入れているといえる。P(X)という述語はそれがああるオブジェクトXに関して真か偽の値をとる「何か」である以外の何者をも意味していない。Xが赤いのか、重いのか、物なのか、人なのか、学生なのか、夫なのか、動いたのか、歩いたのか... 要するに世界に関して何を主張したいのかは一切捨象されているのである。

オントロジーは述語論理表現に必要な「語彙」、すなわち具体的な述語を規定する。世界知識を表すときに必要となる述語を階層的に分類して、各カテゴリに属する概念(述語)が、真理値をとる以外にどのような意味や制約を持つかを規定する。たとえば、tall(X), human(X), student(X), move(X) などの4つの述語は本質的な相違点を持っていること、たとえば、tall は属性であり identity を持たない、human はXの本質的的属性を表す概念であり identity を持つこと、student は role 概念であってXの本質的属性を表さず、プロセスを表す move と同様一過性のものである、などである。そして、これらの異なった述語はそれらが属するカテゴリーに固有の性質を持ち、それに依存する様々な制約や関係を持っている。オントロジーはそれらを明示化し、形式的に厳密に記述する。すなわち「内容の理論」を提供する。

物理学の進歩は、上述の様な数学上の概念と演算・関係を使用しながら、運動学、熱力学、電磁気学、などのように個別の分野で知識を体系化しつつ、各種の理論の統合も研究してきた成果を「蓄積」してきたことにある。一方、知識処理の研究者は「内容はあまりにも領域個性が高い」ということを理由に「蓄積する」ことを放棄してきた。オントロジーに現れる概念階層は、どの抽象レベルまでは「共通の内容の理論」として受け入れることができるかと言う議論も含めて、内容の理論を構築し、蓄積することに貢献する。我々は、数学の様に数少ない概念と操作だけで閉じることはないことを認識しつつ、「知識(内容)の理論」構築へ向けて長期的視野を持って取り組まなければならない。

5.2 オントロジー工学がカバーする範囲は?

完全ではないが、少なくとも下記の課題を研究することは必須であろう。

基幹研究

- ・哲学(存在論)
- ・科学哲学
- ・知識表現

オントロジー設計

常識ベース研究

オントロジー構築方法論

- ・記述言語, 方法論, 構築支援環境
- ・オントロジー比較, オントロジー統合

評価方式

標準化問題

データ/知識変換

知識共有/再利用方式

メディア

- ・メディアオントロジー

- ・知識メディア共通オントロジー
- ・メディア統合

5.3 オントロジー工学に方法論はあるのか？

オントロジー構築の方法論は上述のように提案が行われつつあるが、オントロジー工学となるとおぼつかない状況である。オントロジー工学の方法論はこれから周知を集めて走りながら考えて行くべき物であると考えている。あえて、筆者のこれまでの経験を下にした非常に個人的な見解を述べるならば、

- (1) 従来当然と思っていたことを詳細に検討すること
- (2) 暗黙の仮定を洗い出すこと
- (3) 概念、特に基本概念の定義の明確化、定式化を試みること
- (4) 概念間の関係に注目すると同時に、それがどのような関係であるかを明確に概念化し、定式化すること
- (5) システムの処理自身も概念化の対象にすること
- (6) 明示化した仮定や概念、関係を用いてシステムを記述してみる
- (7) 等々

5.3.1 暗黙の仮定とは？

我々が何かを表明するときには必ず暗黙に仮定しているものが存在する。その典型が対象の「モデル」を構築するときである。ある対象のモデルを構築する際には(1)対象を利用する目的に依存して定まる視点があり、(2)それに強く依存する仮定があり、そして(3)視点に依存して切り出される対象を構成する基本概念とそれらの間の関係がある。ところが、多くの場合実際に記述されるのは「モデル」のみであり、基本概念を除いて全てが暗黙的であり、記述されることはない。しかも、その基本概念も「ラベル」が存在するだけで定義がなされることも多くないのが現状である。そして更に悪いことに、当然合意があると思っていた一般的な用語、たとえば、「モデル」、「振る舞い」、「機能」などの用語の解釈にすら身近の研究者の間で異なった解釈をしていることがある。このように考えると寒気さえ覚えるほどモデル記述は不完全なのである。

5.3.2 具体的な例はないのか？

既知と思っている概念の厳密な定義を試みると、意外と我々はよく分かっていないことに気づいて愕然とすることがある。筆者らが最近行った、「故障のオントロジー」の研究の例を簡単に紹介する[来村 98]。

「故障原因」とはなにか？これは定義する必要性があるとは思えないほど自明に思われるかもしれないが、実は結構曖昧である。テレビが壊れたので修理してもらった時を考える。そこで、修理屋さんに「このコンデンサが壊れてましたので交換しておきました」と言われると、コンデンサの故障が原因であったと思う。しかし、そのテレビは去年買ったばかりであるとする、コンデンサが壊れたことの原因があるように思われてくる。どこかに断線があって定格を超える電流が流れたのではないのか？隣の部品が発熱しているのではないのか？などである。この2種類の「原因」はどのように概念化されるのか？原因から兆候が現れるのであるが、その伝播過程や理解過程は上述の二つの原因では相違があるのであろうか？前者の場合支配的であるのは、「全体(テレビ)」が壊れたのだからその「部分(部品)」のどれかが壊れているはずであるという理解であり、後者の場合にはそのような推論は無効で、物理的現象に基づく因果的理解が必要である。また、実際の因果伝播を考えても、前者の場合にはテレビの構造と部品の機能の知識だけで因果連鎖

を追っていけるが、後者の場合にはそれを越えた知識が必要とされる。因果連鎖の途中には、兆候を説明する物理状態が複数個現れるがそれぞれをどのように概念化すればいいのか？故障の原因に「究極の原因」というのがあるのか？「兆候を説明する何か」を故障原因と呼ぶなら、究極の原因に至るまで何レベルの原因があるのか？どのタイプの原因を発見するにはどのような知識の準備が必要なのか？これまでに多くの故障診断システムが提案されてきたが、それらの診断能力を概念的に比較検討するために十分な語彙(概念)があるのであろうか？....

このように、「故障」というありふれた概念に関してても非常にたくさんの疑問がわいてくる。そして、これらの疑問に答えるために概念化を精密にした後に故障診断システムを再考察すると、多くのシステムの能力の相互比較ができるだけでなく、新しい能力を持つ診断システムの設計が可能になった[来村 98]。

5.4 オントロジー工学はモデルの多様性にどのように対処するのか？

筆者は以前、オントロジー工学の究極の目的は、「実世界、厳密には情報科学が対象とし得る全ての対象のモデル構築の基盤を与えること」にあることを宣言した。しかし、この言明はきわめて一般的であるため補足説明が必要であるように思われる。

問題は「対象」という概念である。情報科学が対象とするものは実に多様である。故障診断の対象を例に取ると、モータ、テレビ、自動車、発電(化学)プラント、日本全体の電話網、というようにその「領域」も「規模」も多様である。しかも実際には診断以外の様々なタスク(問題解決)に対応するモデルを構築する必要がある。このような多様性に対応できるモデル構築の基礎としてオントロジーは機能しなければならない。

上の例は全て「具体的な人工物である」。我々は「問題解決」のモデルにも興味がある。診断とはどういう行為か？設計とは？あるいは問題解決ではないが、ビジネスプロセスもモデル構築の対象となっている。

そして最大の「対象」は「知識」である。知識のモデルとは何か？この問いは興味深いがこのここでは触れないでおこう。

モデルの構築には、基礎となるビルディングブロックとそれらの接続や組み合わせを規定する規約や理論が必要となる。このことは全てのモデル構築に共通する。確かに具体的なビルディングブロックは領域に依存する。しかし、一見多様に見えるモデルも比較的少数のタイプに類型化することができるものと思われる。オントロジーとしてはそれらのタイプに固有の理論を提供することが目的であること、そして、各クラスに属する概念の抽象階層を注意深く構築すれば多くの共通する性質(理論、規約)を見いだすことができることを考慮すれば、それほど大きな困難はない。しかも、オントロジー工学はオントロジー開発の方法論、その記述言語、オントロジーの利用法などを提供してモデル構築という一般的な側面について貢献する。

考察すべき問題は多様性の問題より、モデルに要求する計算のクラスであるように思われる。言い換えると、上述のオントロジーの3つのレベルの考え方である。レベル3までのオントロジーは確かにドメインの多様性に耐えうるものの構築は容易ではない。しかし、レベル2オントロジーは使用に耐えるものの構築の可能性は十分高いと考えている。

6. むすび

本稿では、オントロジーに関する包括的な議論と、海

外の動向を紹介すると共に、オントロジー工学について解説した。最後に重要と思われる課題をもう一度論じてみたい。

(1) オントロジーの標準化

すでに ANSI における活動は紹介したが我が国における標準化の意識は高くないことに危惧を覚える。標準化という言葉の Negative な側面を強く意識しすぎる国民性が原因かもしれない。知識の共有や interoperability, ドキュメントの検索や再利用などは基本語彙や概念, そして最小限のタグなどを共通化する以外に効果的な方法はないことはほとんど自明ではなからうか。そもそも人間がコミュニケーション可能であるのは語彙を共有しているからである。我が国からオントロジーの分野で標準化案を提案する位の意気込みと実質的な活動が生まれることを期待したい。

(2) オントロジーに関する共通の理解

既に述べたように、未だにオントロジーの定義については研究者の間で合意は得られていない。というより、むしろ"オントロジーとは何かという議論はやめよう"という動きがでてきている状況にある。少なくとも、自然言語処理に携わるもの、知識ベースの立場をとるもの、哲学的な立場をとるものの3者では合意を得ることは困難であるように見受けられる。自然言語研究者は"語彙"あるいは"辞書"という観点を重視し、知識ベース研究者は"実行可能性"や知識ベース構築の"部品"という考えを重視し、哲学者は"共通性、汎用性の高い上位モデル"の立場を固執する。いずれも十分強固な正当性の根拠を持っており議論を収束する方向に持っていくことは容易ではない様に見受けられる。

(3) オントロジー工学への道

このような状況ではあるが、オントロジーが重要であるという点では研究者の合意は強固である。オントロジーの厳密な定義には触れないようにして、各自が必要と考えるオントロジー理論、システム、そしてオントロジー自身を開発する事に専念することで今は十分であるようにも思われる。現在はオントロジー研究は幼児期にあり、はじめの一步で躓かないことの方が重要であろう。

そこで重要なことはオントロジーの定義ではなく、オントロジー工学の旗印の下で何をなすべきかということをも明確にすることでと考える。そして、過度な期待は慎みつつ、野心的で挑戦的であることが大切であろう。

Researchers be reasonably ambitious!

オントロジーの問題は多くの人が「できるかもしれない」と positive に考え出すことが必要である。その前向きな姿勢が困難を切り開いて新しい世界を作り出していくことであろう。

謝辞：いつも激論を戦わしている、大阪大学産業科学研究所、池田助教と来村助手に感謝します。JIPDEC オントロジー工学委員会での討論も大変有意義であった。ここに委員の方々に謝意を表します。

<文献>

- [Breuker94] J.Breuker and W.V.de Velde: The Common KADS Library for Expertise Modelling, IOS Press, Amsterdam, 1994.
[Cyc] <http://www.cyc.com/public.html>
[EWN] <http://www.let.uva.nl/~ewn/>
[DC97] Dublin core metadata;
http://purl.org/metadata/dublin_core, 1997
[Guarino 97] N.Guarino: Some organizing principles for a unified

top-level ontology, Working Notes of AAAI Spring Symposium on Ontological Engineering, Stanford, 1997.

[GUM] <http://www.darmstadt.gmd.de/publish/komet/gen-um/newUM.html>

[IDEF5] <http://www.kbsi.com/idef/idef5.html>

[Ikeda97] M. Ikeda, K.Seta et al., Task Ontology Makes It Easier To Use Authoring Tools, Proc. of IJCAI-97, pp.342-347, 1997.

[池田98] 池田満: 知識工学の基盤としてのオントロジー, 人工知能学会誌, Vol.13, No.1, pp.11-13, 1996.

[IMS97] (IMS9717)知識の体系化:設計及び製造のための構築システムに関する研究成果報告書(普及版)製造化学擬態センタ, IMS センター, 1998.

[JIPDEC98] 大規模知識ベースに関する調査研究—オントロジー工学に関する調査研究—, 日本情報処理開発協会, 1998.

[来村98] 来村他:故障オントロジーの構築:その利用と効果, 人工知能学会研究会資料 SIG-IES-9703-5, 1998.

[久保98] 久保他:オントロジー構築ガイドシステムの概念設計, 人工知能学会全国大会(第12回), 1998.

[MCF97] Meta Content Framework using XML; <http://www.w3.org/TR/NOTE-MCF-XML-970624>, 1997.

[Miller93] G.Miller, et al.: Five papers on WordNet, CSL Report 43, Cognitive Science Laboratory, Princeton University, 1993.

[溝口96] 溝口理一郎, 形式と内容—内容指向人工知能研究の勧め—, 人工知能学会誌, Vol.11, No.1, pp.50-59, 1996.

[Mizoguchi95] R. Mizoguchi, et al.: Ontology for Modeling the World from Problem Solving Perspectives, Proc. of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, pp.1-12, 1995

[溝口97] 溝口理一郎,オントロジー工学序説—内容指向研究の基盤技術と理論の確立を目指して—,人工知能学会誌, Vol.12, No.4, pp.559-569, 1997

[Ogino et al.97] Takano Ogino, et al.: An Experiment on Matching EDR Concept Classification Dictionary with WordNet, IJCAI-97 Workshop on Ontologies and Multilingual NLP, pp.23-27(1997)

[RDF97] Introduction to RDF Metadata; <http://www.w3.org/TR/NOTE-rdf-simple-intro-971113.html>, 1997.

[瀬田98] 瀬田和久, 他:問題解決オントロジーの構成—スケジューリングタスクオントロジーを例にして—, 人工知能学会誌, Vol.13, No.4, 1998 (掲載予定).

[Sowa95] J.Sowa: Distinction, combination, and constraints, Proc. of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.

[Sowa98] J.Sowa: Knowledge representation: logical, philosophical and computational foundation, PWS Publishing Comp, Boston (to be published in 1998).

[Takeda95] H.Takeda, et al.: Agent organization and communication with multiple ontologies, International Journal of Cooperative Information Systems, Vol.4, No.4, pp.321-337, 1995.

[TOVE] <http://www.ie.utoronto.ca/EIL/tove/toveont.html>

[ティヘリア93] ティヘリア・ジュリア・A., 他, タスクオントロジーと知識再利用に基づくエキスパートシステム構築方法論, 人工知能学会誌, Vol.8, No.4, pp.476-487(1993年)

[上田97] 上田他:設計意図を考慮した生産工程再設計システムの検討, 人工知能学会全国大会(第11回), 29-01, 1997.

[WordNet] <http://www.cogsci.princeton.edu/~wn/>

[山口98] 山口他:法律オントロジー, 人工知能学会誌, Vol.13, No.2, pp.189-196, 1998.

[Yamaguchi97] T.Yamaguchi, et al.: A legal ontology rapid development environment using a machine-readable dictionary, LEGONT'97, pp.69-73, 1997.

[横井97] 横井俊夫, 仲尾由雄, 荻野孝野, 田中裕一: 概念レベルにおける電子化辞書の情報構造, 情報処理学会論文誌, Vol.38, No.1, pp.32-43(1997).