オントロジーと知識処理

溝口理一郎@大阪大学

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

1.はじめに

エキスパートシステムが表舞台から姿を消して久しい.エキスパートシステム技術が産業界にしっかり根付いて,当たり前の技術になってしまったことも原因の一つであるが,やはり知識ベース技術の,ひいては知識処理技術全般の閉塞感がその大きな原因となっていることは否めない.とは言うものの,「コンピュータに知識を蓄えて何か人あるいは企業の助けになることやらせたい」という欲求を消すことができないのも事実なのである.

本特集の主題である「言葉と知識」から, エキスパートシステムや知識ベースシステム などが昔ながらの格好で再び日の目を見よう としているのではないかと警戒される向きも あるかもしれない.しかし,その心配はご無 用である.この特集は言葉と知識処理との新 たな結びつきを再考して,来るべき情報化社 会に必要となる,と言うよりは情報化社会の 実現に貢献する技術と理論の解説である.

実際,一時の人工知能ブームが沈静化して, 各社が人工知能や知識処理と名の付く部署の 看板の掛け替えを一斉に行った時期に比べる と,知識処理の必要性は高まりつつある.イ ンタネット上の膨大な知識の検索と処理,知 識マネジメントに代表される企業に蓄えられている知識の管理・活用、STEP(製造物データモデルの標準化活動)におけるモデル設計、コンカレントエンジニアリングにおける多様なデータ・モデルの交換・統合、XML タグセットの系統的設計と意味定義、及び XML ドキュメントの知的利用、EDI(電子データ交換)におけるデータモデル設計、インテリジェントエージェント、Web に基づく学習支援システムのインテリジェント化、・・・枚挙にいとまがない・・しかし、これらの要求の多くは従来のエキスパートシステムとはだいぶ様相が異なっている・相違点は以下の通りである・・

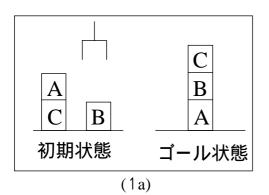
- (1)システムによる自動問題解決を目指すのではなく利用者とのインタラクションが 重視されること
- (2)知識の共通化,標準化を指向すること
- (3) 専門家の経験則ではなく,対象のモデルに関わる基盤知識が対象となること

本稿では,この様な新しい知識処理を根底から支えそれを可能にする理論と技術であるオントロジー研究を紹介する[溝口 99b,99c].

2.オントロジーとは

オントロジーとは本来哲学用語で,「存在論」と言う意味であるが,コンピュータ科学では『情報処理が対象とする世界のモデル構築者がその世界をどのように「眺めたか」,言い換えるとその世界には「何が存在している」と見なしてモデルを構築したかを(共有を指

向して)明示的にしたものであり,その結果 得られた基本概念や概念間の関係を土台にし てモデルを記述する事ができる』と言うよう に理解されている.



概念化1

オブジェクト 関係

block A	on(X, Y)
block B	above(X, Y)
block C	clear(X)
table A	holding(X)
hand A	handEmpty
	(1b)

概念化 2

オブジェクト 関係:

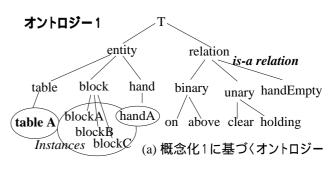
block A	on(X, Y)
block B	above(X, Y)
block C	onTable(X)
hand A	clear(X)
	holding(X)
(1c)	handEmpty

図1 ブロックワールドの概念化

2.1 オントロジーの具体例

図1aのブロックワールドを例にしてオントロジーの具体例を示そう.まず,基礎となる概念は「概念化」である.概念化とは対象とする世界に存在する概念とそれらの間に成立する関係を言う.概念化は対象世界の骨格

となる物である.このブロックワールドでは ロボットハンドを使って与えられた配置の積 み木を目標状態のように積み替えることがタ スクである.そこではその他の一切の作業も,



Axiom: above(X,Z):-on(X,Y), on(Y,Z).

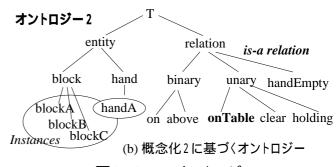


図2 2つのオントロジー

物体も存在しない.この問題の場合,典型的な概念化は図1bに示す様なものである(概念化1).この世界に存在するオブジェクトとしては3つのブロックと,ロボットハンド,そしてテーブルである.そしてそれらの間に存在し得る多くの関係の内,このタスクで興味がある関係は,ブロックや机の間の上下関係(on,above)そしてロボットハンドがブロックをつかんでいるかどうか(holding,handEmpty)とブロックの上に何も乗っていないかどうか(clear)だけである.それ以外の関係は一切不要である.

ここで重要なことはこの概念化1以外にも意味のある概念化があることである.それを図1cに示す(概念化2).相違点はテーブルaの存在をオブジェクトとして認識するか,関係として認識するかである.少し考えてみれば,この世界ではテーブルaをオブジェクトとして操作の対象にする必要は無く,状態としてブロックが直接机の上にあるかどうかさしてブロックが直接れの上にあるかどうからにonTable(X)という関係としてテーブルaを概念化する方法があり得るわけである.そしてこの方が概念化1よりこのタスクの実行が簡単になる可能性がある.これらの概念化から作られるオントロジーを図2に示す.

さて、問題はこのタスクを実行するプログ ラム作成なのであるが,通常プログラムを作 成しても残っているドキュメントは機能仕様 や詳細設計くらいで、対象となった世界をど のように眺めたか、すなわちそのプログラム が深く依存している「概念化」が明示的に記 述されることはほとんど無い.そして,この 様な二つの異なる概念化に基づいて作られた プログラム同士は,もし協調して作業をする ような会話をする場合があったとすると話が 通じない.両者はテーブル a の存在に関して 認識を共有していないからである.これはプ ログラムの開発者同士が会話をしたとしても 同様である.更に悪いことは,概念化1はタ スクに依存しない,かなり一般的な概念化で あるが,通常のソフトウェアがそうであるよ

うに,概念化2の方はタスクに依存して簡略 化してある所に問題があり,開発に携わって いない人と開発者が会話をするとんでもない 誤解の原因となり,それに当分気づかないと 言う事態が起こり得るである.

2.2 トップレベルオントロジー

上述のオントロジーはごく普通のオントロジーであるが,ここでは現在筆者らが研究室で設計中のトップレベルオントロジーの上位のカテゴリを図3に示す.まだ完成していないが,感触だけはつかんでいただけると思う.

- -基底 -空間 -時間 -物質 -実在物 -具体物 -物 - プロセス -行為 -現象 -抽象物 -関係 -ロール -性質 -属性 -単位 -出来事 -事実 - 状態 - 値 -表現 -表現形態
- 図3 トップレベルオントロジーの例

紙面の余裕がないので簡単に説明する.上

の二つ(基底と実在物)はこの世に存在する物を表すカテゴリーで、それ以下はそれらを説明したり、それらについて述べるために必要なカテゴリーである。このオントロジーでは実在物が本質的に依存する「基底」を実在物から明確に分離している。

「プロセス」の下位の「行為」と「現象」 の相違点は,前者には行為者がいて,後者に は行為者がおらず,単なる「参加者」がいる だけであることである. そしてコンピュータ モデルで起こっていることも含めて現実世界 で起こっていること(=プロセスのインスタ ンスや物のインスタンス)とそれを認知して 記述した物(=事(出来事,事実,状態など)) を分離している.その結果,上には表現され ていないが、「表現」は「内容」としての「事」 と「表現形態」とから構成されると定義され ている.このトップレベルオントロジーの優 秀性は今は不明であるが,この様なものを持 つことによって通常のオントロジーを設計す る際の指針を容易に得ることができる.実際, ロール(学生,妻,医者,歩行者など)は「物」 の下位ではなく別の概念としてトップレベル の地位を与えてあり,よく起こりがちである 概念の混同を避けることに貢献する.

哲学のオントロジーは「存在一般」を対象として、それを説明する「カテゴリ」の体系の設計を目指している。コンピュータ科学でいうオントロジーはこの様なトップレベルオントロジーと呼ばれるオントロジー以外は、

存在そのものを対象にすることはなく,開発者が興味を持つ任意の世界(対象物)を対象としてオントロジーを設計する.この様に対象は哲学とは異なるがその方法論と精神は哲学におけるオントロジー設計と同様である.このことは図4を例にして後で述べる.

3.何はオントロジーではないか?

余り厳密にオントロジーを規定する事は, 排他的になり関連研究を抑制する悪影響があ るので必ずしも好ましくないが,かといって 「何でもかんでもオントロジー」と言ってい るとこれまた見識のある方々から「昔からあ る概念に新しい名前を付けて人心を惑わす な」というお叱りを受けると言うジレンマが ある.ここでは,後者に気を使うことにして オントロジーと言い難いものを明示化して, 多少堅めの態度をとってみようと思う.

・オントロジーは単なる語彙集合ではない

オントロジーは共通語彙を提供する.しかし,語彙がそのままでオントロジーとなるわけではない.まず,語彙と概念の区別が重要である.語彙(用語)は概念を表現するための「ラベル」としての機能を持っており,自然言語依存であり,同義語や別名などの問題に敏感である.オントロジーは「語彙」ではなく「概念」の体系である.

・オントロジーは単なる概念階層ではない

オントロジーは概念分類(Taxonomy)を包含する.しかし,概念分類それ自体がオントロ

乗り物 乗り物(世界) -種類 - 陸上車 -陸上車 - 車 -船 - 4輪車 -飛行機 - 乗用車 -機能 - トラック -人を運ぶ - オートバイ -荷物を運ぶ - 汽車 -属性 -馬力 - 船 - 大きさ - 飛行機 -構成物 -エンジン (a)単純な概念階層 -車体 -交通システム

(b) オントロジー

図4 乗り物オントロジー

ジーとなることはない . 例を挙げよう . 「乗り 物オントロジー」を設計している場合を考え てみる .図 4 (a) に標準的な Taxonomy を示す. 乗り物には陸上を走るもの,水の上を走る 「船」, 空中を飛ぶ「飛行機」に分かれる.そ して,更に陸上車は「乗用車」と「トラック」 に分かれると言うように分類される.しかし, この分類階層を見て、「乗り物の事が分かっ た」と言えるであろうか?確かに乗り物の種 類は分かったが何か物足りない.そこで,図 4(b)の様な階層を考えてみる.これはいわゆ るトップレベルオントロジーを参考にして作 ったものである.乗り物を説明するには,種 類だけではなく「機能」「属性」「構成物」そ してそれが使用される環境としての「交通シ ステム」などが本質的に必要で、それらは乗 り物世界のトップレベルカテゴリとして認定 すべき物であるように思われる..トップレベ ルにある概念と最上位の概念との関係は,

is-a と言うよりは part-of に近い関係にあるといえる.これはちょうど哲学におけるオントロジーが,「存在」を説明するために必要十分なカテゴリーは何か?と言うことを考察していることと類似している.

・オントロジーは知識表現ではない

オントロジーは意味ネットワークでもないし、フレームでもない・内容指向研究を支えるオントロジーは表現自体にではなく「表現対象」に興味がある・オントロジーは対象をモデル化するためのガイドラインを提供する・そしてそのために、オントロジーとして何をトップレベル概念に据えるか、「もの」を構成する部分の同定、同定された概念間の関係の同定等が最重要事項となる・知識表現言語が必要となるのはそれらの基本的な考察の結果、すなわちオントロジーを表現するためである・

・オントロジーは知識ベースと同じではない

確かにコンピュータに何かを記述して蓄積 すればどんな物でも知識ベースと呼べる.し かしその考えは不毛である.オントロジーは ある知識ベースが前提としている対象世界の 概念化を明示化した物であり,オントロジー は知識ベースとはメタな関係にある.

4 . オントロジーの役割

4.1 オントロジーの効用

オントロジーの知識処理に対する効用は以 下のようにまとめることができる.

(0) 合意を得る手段

他人と合意を得る事は容易ではない. その 原因は沢山あろうが、一つには「知識」に関 する合意を得ようとすることにあること、言 い換えれば細部に渡る合意を得ようとするこ とにある、オントロジーは「知識」そのもの ではなく、「知識」の前提となっている概念化 に関するものであり,いわば対象世界の「骨 格」を明示化した物である.「これに合意無く して協調は不可能」とも言える根元的な物な のである.複数の人が合意する事が出きる物 を探すという「合意」の下にお互いのオント ロジーを出し合って、それを題材にして議論 をして共通のオントロジーを設計していき、 最終的に合意できた内容としてのオントロジ ーを得る.合意のレベルが低いとオントロジ ーは非常に抽象度の高い物となり,深い合意 が得られればオントロジーの概念階層も深く なり、詳細な概念構造にまで及ぶものが得ら れる.この過程で遭遇する最大の困難は共通 の対象世界を眺める「観点」の相違である. できるだけ「観点(多くの場合は興味あるタ スク)」 独立に対象世界を眺める事であり,自 分が採用している「観点」をいつも明示する ことを心がけることである.この様にオント ロジーは合意の結果を表すだけでなくそれを 得るための手段として用いることができる. この場合のオントロジーは以下に述べるレベ ル 1 オントロジーで十分である . .

(1) 暗黙情報を明示化

通常無意識の内に仮定したり、前提として

いる概念を明示化する.そのようなものの代表が対象世界の「概念化」である.知識ベースはもちろん,一般にソフトウェアは何らかの概念化に基づいているが,既に述べたように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.

(3) 再利用と共有

知識の共有と再利用は困難である事は論を 待たない、その理由は沢山あるが、大きな要 因の一つには仮定されている前提が暗黙的で あることにある.そして,通常知識と呼ばれ る専門家の経験則などは多様な基本的概念の 複合体であり、問題解決コンテキストへの依 存性や主観的要素が強いものである.これで は共有や再利用が困難であるのは自明である う.しかし,オントロジーは違う.知識を構 成する基本概念に立ち戻って、知識の元にな る対象世界を客観的な存在として考察するこ とによってそのような知識を構成する基本概 念を同定する、そして知識の抽象度に応じた 階層性,知識の分解可能性,そしてコンテキ スト(特にタスク)依存概念の同定と除去な 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, 共有・再 利用可能な知識を見出す糸口を与える.

(4) 知識の体系化(コンピュータ上での)

人間はこれまで文字と本を使った知識の体 系化を行ってきた.しかし,それは人間のた めの体系化でありコンピュータには理解できない.コンピュータ上で知識の体系化ができるとすればその有用性は極めて高い事は言うまでもないであろう.さて,知識の体系化にとって最も重要なことは,関係する対象世界を支配する概念を明確化することと知識を記述するための共通語彙を定めることであるが,オントロジーはその両方を与える.そして言うまでもなくそれはコンピュータが処理可能である.

(5) 標準化

オントロジーは少なくともあるコミュニティーで共有されることを目指して開発されるが、そのことが示唆するように、オントロジーに含まれる語彙と概念は共通性が高く、それは標準化への本質的な第一歩となる。

(6) メタモデル的機能

「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)」と言うオントロジーの定義とオントロジーをクラス定義と見なして、そのインスタンスを生成しながらモデルを構築するオントロジー利用プロセスを考えれば、オントロジーはモデル構築に必要な基本概念とガイドラインを提供する機能があることが分かるであろう.

(7) 総合的効用

これまで述べてきたオントロジーが持つ効用を眺めてみると,オントロジーが如何に有用であるかが見て取れる.実際,人々の合意

形成に役立って,通常暗黙となっている基本的な世界観(概念化)が明示化され,それが人々に共有されると同時に人々が持つ知識の根元となる概念が明示化され,それらを標準化して(少なくとも企業内で共通化して)必要であればそれらを用いて知識を再記述して体系化を行うと同時に必要なモデル構築を行う.もちろん,そのようにして開発されたモデルとなる.

多少楽観的すぎる表現ではあるが,オントロジーは上述のことを実現する可能性を持っている事は確かである.このことを考慮すればオントロジーは知識マネージメントや XML 文書タグ設計の基礎,及びそれらの有効利用に大きく貢献することも理解されよう.

4.2 オントロジーの3レベル

少し技術的な事に立ち入るがオントロジー の計算的な意味の深さについて触れておこう.

・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.前述のように,このレベルのオントロジーは主に人間用であり,自然言語を用いた定義でも十分機能する.

・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の際に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.これをレベル2オントロジーと呼ぶ.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に基づく問題解決における実行のパフォーマンス,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できる.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.KADS のオントロジー[Breuker94]やタスクオントロジー[テヘリノ93][瀬田98]はこのレベルのオントロジーである.

5.分野毎のオントロジー

オントロジーの開発例としてはStanford大学のオントロジー記述言語Ontolinguaを使って開発されたオントロジーベースがあげられるが[Ontolingua],ここでは,それとは別にこれまでに開発されたオントロジーをいくつ

か紹介しよう.

5.1 法律オントロジー

静岡大学の山口等による法律オントロジーの開発は独自の構築支援ツール Dodd Ie を法律の専門家が利用して構築された物である点が特長である[山口 98,99].

5.2 ビジネスプロセスオントロジー

ビジネスプロセスモデリングやエンタプライズモデリングの分野でのオントロジー開発も盛んである[伊藤 98].特に,エディンバラ大学の AIAI 研究所[AIAI]とトロント大学[TOVE]で開発されたエンタプライズオントロジーは特筆に値する.共に,開発の過程に関する論文もあり参考になる[Uschold 98][Gruninger 95].

オントロジーの情報検索における概念階層 の利用は典型的な応用例であるが,実際その 利用例は多数ある.

5.3 設計におけるオントロジー

設計にかかわる世界を対象として機能概念を中心としたオントロジー設計が行われている.機能概念にまつわる3種類のカテゴリ(ベース機能,機能タイプ,メタ機能)の同定,各カテゴリ内での機能概念の階層的組織化,ベース機能概念の振る舞い・構造概念へのGrounding,そして対象の機能理解システムの開発などを行い,包括的な機能オントロジーの開発と利用,そしてそれに基づく設計知識の体系化研究が行われている[溝口 99a].

東京大学の冨山等も長年の知的 CAD の研究

の成果である KIEF のオントロジーを抽出して 見通しの良いシステムの記述・構築を行って いる[関谷 99].

5.4 故障オントロジー

故障というありふれた概念の概念化を精密にした後に故障診断システムを再考察した結果,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 99a, 99b].

5.5 辞書とオントロジー

機械可読辞書の開発が我が国も含めて活発に行われている.EDR の辞書, WordNet, EuroWordNet などである.いずれも上位の構造はオントロジーに対応する.

5.6 その他

学習支援システムの分野では極めて活発にオントロジー研究が行われている.特に教材開発のためのオーサリングツールにはオントロジーのメタモデル機能が本質的であり,オントロジーベースのオーサリングツールの研究が活発である[Ikeda97][Mizoguchi,99d].奈良先端大の武田等は知識コミュニティーとオントロジーとの関わりの研究を行っている[武田 99]

6.これからの知識処理

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[溝口 96]に関するパラダイムシフトである.これまでの AI 研究では一般的な議論が可能な「入れ物」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重視されて来なかった.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たのである.

これまで述べてきたようにオントロジーは 内容を扱う理論と技術を提供する.オントロジー工学は知識の内容,そして知識の源である「対象世界の成り立ち」を考察する.そして,部分的にせよコンピュータが理解可能な形で基本概念を構造化して,現実の対象をモデル化する際のビルディングブロックを提供し,モデル化の規約を与える.

一方,マルチメディア処理の本質は明らかに内容を理解することを通したメディア統合にある.多種多様なメディア、すなわちテキスト、音声、図、静止画像、動画像等によって表現されている知識を、各種メディア間での変換、要約、統合等を行う基盤技術が開発されなければならない.ここでもオントロジーが重要となる.メディア自体のオントロジーとメディアの「内容」に踏み込んだオント

ロジーとを用いた「理解によるメディア統合」 は緊急の課題と言える.

知識ベースシステムが独立した自動問題解決器であることを止めて、上述のパラダイムシフトに順応し、オントロジー工学で武装して生まれ変わりつつある.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)に基づく知識処理は新しい技術として今後ますます重要となろう.

参考文献

[AIAI] http://www.aiai.ed.ac.uk/ entprise/enterprise/ontology.html

[Breuker 94] J.Breuker and W.V.de Velde: The Common KADS Library for Expertise Modelling, IOS Press, Amsterdam, 1994.

[Gruninger 95] Gruninger, M., and Fox, M.S., Methodology for the Design and Evaluation of Ontologies, Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal, 1995.

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

[伊藤 98] 伊藤,山口:オントロジーエンタープライズモデル,人工知能学会誌, Vol.13, No.6, pp.870-879, 1998.

[来村 99a] 来村他: 故障オントロジー - 概念抽出とその組織化 -, 人工知能学会誌., Vol.14, No.5, 1999. [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ー,人工知能学会誌 Vol.11,No.1,pp.50-59(1996)

[溝口 97] 溝口理一郎,池田満: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一 人工知能学会誌 Vol.12,No.4,pp.559-569(1997)

[溝口 99a] 溝口理一郎,来村徳信:設計とオントロジー, 日本学術会議50周年記念シンポジウム「設計の質的転換」,pp.1-8,1999.

[溝口 99b] 溝口:オントロジー研究の基礎と応用, 人工知能学会誌, Vol.14, No.6, pp., 1999[溝口 99c] 溝口,池田,来村:オントロジー基礎論, 人工知能学会誌, Vol.14, No.6, pp., 1999.

[Mizoguchi 99d] Mizoguchi, R.: Ontology-awareness for intelligent instructional systems, Proc. of ICCE99, IOS Press, pp. ,1999.

[Ontolingua] http://www-ksl-svc.stanford.edu:5915/ [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, pp.597-608, 1998. [武田 99] 武田英明,西田豊明,知識コミュニティープ ロジェクト(第6報),人工知能学会全国大会(第13

[ティヘリノ93] ティヘリノ・ジュリ・A.,他、タスクオントロジーと知識再利用に基づ〈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)

[TOVE]

回),25-04, 1999.

http://www.ie.utoronto.ca/EIL/tove/toveont.html

[Uschold 98] Mike Uschold, Martin King, Stuart Moralee and Yannis Zorgios, The Enterprise Ontology, The Knowledge Engineering Review , Vol. 13, Special Issue on Putting Ontologies to Use (eds. Mike Uschold and Austin Tate). (Also available from AIAI as AIAI-TR-195) 1998.

[山口 99] 山口高平他,計算機可読辞書を利用した領域オントロジー構築支援環境,人工知能学会誌,Vol.14, No.6, 1999.