

概念化アスペクトに基づくオントロジー構築

About Ontology Building based on Conceptualizational Aspect

上田 俊夫
Toshio UEDA

池田 満
Mitsuru IKEDA

北陸先端科学技術大学院大学 知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST)

This paper describes the ontology building support based on Conceptualizational Aspect (CA). CA conceptualizes the aspect of concept, considering both human readable representation and machine processable representation, which clarifies ontology expert's know-how to discriminate concepts and relations for ontology's reusability. CA can lead to support for ontology building and using.

1. はじめに

オントロジーは対象世界における暗黙の仮定を概念体系として明示化したものであり、その世界における人間の合意内容を表している。適切に構築されたオントロジーを規約として知識を記述することによって、それらの知識はオントロジーの下で一貫性を有し、その世界の合意を反映したものとなる。そのためオントロジーは知識を積み上げる共通基盤として機能することが期待される。近年では医療、ナノテク、法律、教育など様々な分野で共通オントロジーを構築して、そのオントロジーに基づいた知識管理を進める動きが活発化してきている。

構築されたオントロジーが知識を積み上げる共有基盤として適切に機能するためには、特定の問題に依存したものではなく、再利用性を考慮されたものである必要がある。そうしたことに考慮してオントロジーを構築・利用できる能力をオントロジーリテラシー(OL)と呼ぶ。しかしながら OL はオントロジーの構築・利用経験に基づいて獲得されるものであり、その習得は容易ではないことからオントロジー構築の敷居をあげてしまっている普及の障害となる。そのため、オントロジー構築対象世界の内容を理解できるのは人間であるオーサしかいないという立場に立てば、OL を補充して、オーサの理解した内容について概念の峻別を手助けすることが必要であり、そのために人間が理解しやすい表現と機械が処理しやすい概念化した表現の両方を対応づけて扱えることが求められる。

そこで我々は、人間が概念の共通の性質をまとめて捉えようとする対象の捉え方に注目し、人間がどのように対象を捉えているかを表現しやすいフレーム表現と記号処理可能な表現を対応づけた**概念化アスペクト(Conceptualizational Aspect : CA)**を媒介として、オントロジー構築・利用を支援するというアプローチを提案する。

CA を用いることで、オントロジーエキスパートの持つ対象の捉え方についての知見を表出し、その知見を CA オブジェクトとして再利用可能な形で蓄積することができ、蓄積された CA オブジェクトを用いることで、オントロジー構築時には概念峻別の手掛かりとなる。さらに既存オントロジーに対しては、CA オブジェクトによって捉え方が明示化され、メンテナンスに効力を発揮するようになる。

CA を中心としたオントロジー構築・利用を進めることで組織

内のオントロジー構築・利用サイクルを回し、組織員の OL 養成にもつながると考えている。

本稿の構成は以下の通りである。2節では、基本となる概念表現を提案し、その概念表現で概念の峻別さえできれば、再利用性を考慮してオントロジーを正規化できることを示す。3節では、概念化アスペクト(CA)について説明し、CAが概念峻別の手掛かりとなることを示す。4節ではCAに基づくオントロジー構築支援の概要について説明し、最後にまとめとする。

2. 概念表現

2.1 オントロジー記述言語としての OWL の問題

我々は、オントロジー構築の基盤をOWLとすることになっている。そのインターオペラビリティやスケラビリティが優れていることがその理由であるが、OWLは、そのまま利用するにはいくつか問題がある。その問題の一つに記述論理推論による振る舞いを持つことで、オーサはその振る舞いとらわれて暗黙的な概念化を行うという混乱が生じるということである。それはW3C自身がサンプル公開しているワインオントロジー¹にさえ垣間見ることができる。ワインオントロジーでは *locatedIn* プロパティを *CaliforniaRegion locatedIn USRegion* のような地域の部分全体関係、*CaliforniaWine locatedIn CaliforniaRegion* のようなワインの生産地の2つの意味で使うことで *CaliforniaWine subClassOf AmericanWine* という包摂関係と *CaliforniaWine locatedIn USRegion* という推移関係を導出している。振る舞いに囚われたより極端な例としてはOWLをプログラミング言語として使い、数独を解くプログラムを作成している例もあり²、OWLは、オントロジーに期待される概念化の暗黙的な仮定を明示することに必ずしも貢献しない使われ方がなされる。

そこで OWL で適切にオントロジーの構築ができるようにオントロジー基礎理論の考察と OWL の言語的特徴を考慮した概念表現を新たに定義し、その概念表現に従って OWL 上でオントロジーを構築することで、正しいオントロジーを構築することを支援する。

2.2 概念表現

オントロジー工学研究の大きな成果の一つにロール概念の峻別がある。ロール概念は、ある状況に依存性して定義される

連絡先: 上田 俊夫, 北陸先端科学技術大学院大学 知識科学研究科 mailto:ueda_t@jaist.ac.jp

¹ <http://www.w3.org/2001/sw/WebOnt/guide-src/wine.owl>

² <http://www.mindswap.org/~aditkal/sudoku>

概念であり、適切に区別しなれば、オントロジーの再利用率を下げてしまう危険性があることが指摘されている。OWLには、ロール概念を明示的に記述するプリミティブは用意されておらず、そのままでは表現できない。

一方、OWLには必要十分条件で定義されたクラス(定義クラス, **defined class**)を用いて、あるクラスが定義クラスに包摂されるかを導出したり、ある個体が定義クラスのインスタンスであるかを判定したりできる。この仕組みを応用すれば、状況に依存する概念を定義クラスとして定義しておき、個体の状態が変化したときに、その概念への所属性を推論することによって、個体がその状況に依存しているかを判断することができる。

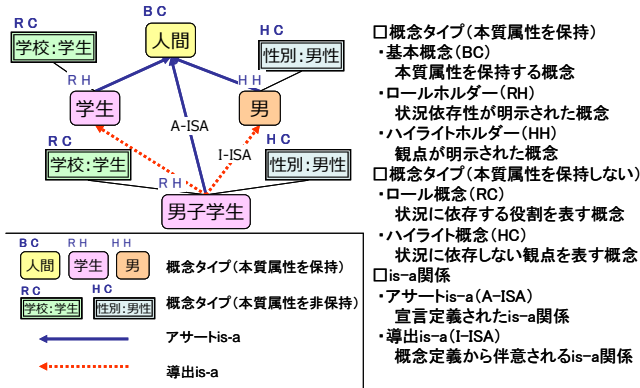


図 1 概念表現とその表現例

この理解を基に、定義クラスによる伴意をオントロジー基礎論の考察を反映させることで概念表現を定義した。その概要を図 1 に示す。

この概念表現では、概念の型を表す概念タイプが 5 種類、is-a 関係は 2 種類が定義されており、オントロジー基礎理論で定義される本質属性の分析から、各概念タイプ間を満たすべき制約を規定している。例えば、個体は、本質属性をただひとつしか保持できないので、本質属性を継承するアサートis-aはただ一つだけ持つことや、状況に依存しない基本概念は、状況依存性のあるロールホルダーの下位概念にはなれないなどが規定されている。概念表現の詳細については[上田 06]を参照のこと。

なお、この概念表現では、is-a 関係以外の関係、属性、制約は明示されない。これは、関係、属性、制約を扱わないという意味ではなく、本研究が扱う概念階層レベルの構築支援では、抽象化しているだけである。後ほど説明する概念化アスペクト(CA)では、その種類によって異なる概念間の制約が定義されており、OWL にシリアライズする段階で関係、属性、制約などが顕在化される。

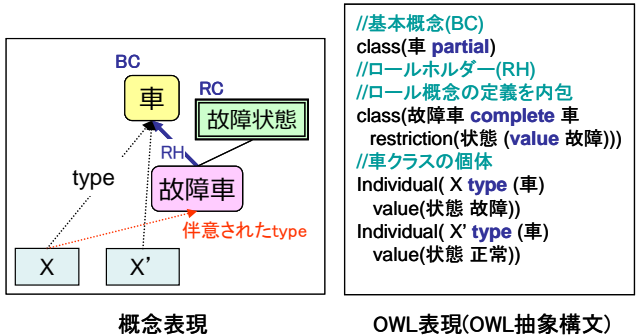


図 2 概念表現と OWL 表現の対応

なお、提案する概念表現では、図 2 のように OWL 表現と対応付けられる。OWL 表現ではロール概念が明示される訳ではない。図 2 の OWL 表現では、車のインスタンスである個体「X」は、その状態が故障であるので定義クラス「故障車」に伴意されるが、状態が正常の「X'」は伴意されることはない。このように OWL 表現では、状況依存性を伴意によって表現し、概念表現における「故障状態」がロール概念であることや「故障車」がロールホルダーであることは明示的に表現されない。

2.3 オントロジーの正規化

提案する概念表現は、再利用率が考慮されたオントロジーが満たすべき制約が表現されている。その制約を使うことでオーサの直感的理解を表す概念階層を入力し、その階層内に問題点を発見して、再利用率が考慮されたオントロジーとして正規化されたものを出力することが期待できる。

[上田 06]では、オントロジーの正規化を実現するために「①問題点を発見するための知識」、「②問題理由を指摘するための知識」、「③問題理由に基づいて修正案を提示する知識」の 3 種類の知識を用意し、それらの知識を用いてオントロジーを正規化する方法を提案している。それらの知識を使うことで、例えば、①多重継承がされている概念がある場合は、②その本質属性が曖昧になることを問題点として指摘し、その場合は、③その概念のアスペクト概念が切り出せないかをオーサに問い合わせ、切り出せる場合は、導出is-aを用いて表現することを提案するといった支援が可能となる。

本項では概念表現を提案し、概念表現に従って概念の峻別ができた場合は、再利用率を考慮してオントロジーの正規化を行えることを示した。しかしながらこのままでは概念の峻別自体、オーサのオントロジーリテラシーに依存しているため、その支援が必要となる。次節では、概念の峻別の手掛かりとなる概念化アスペクトを説明する。

3. 概念化アスペクト

3.1 概念峻別の支援

本節では、オントロジー構築支援の従来研究が概念峻別についてどのようなアプローチを取っているのかを比較することから、我々の研究の位置づけを説明する。

(1) 溝口のアプローチ

溝口[溝口 05]は、知的システムを構成するために内容を扱わなければならないが、内容がどんなものかが分かるのは人間だけであり、結局、人間が概念を峻別しなければならないという立場を取る。そこで人間が概念を表現し、峻別しやすいグラフィカルな表現を提供し、その峻別をどのように行えばいいのかについて必要なガイドラインを本[溝口 06]として提供している。

(2) A. Rector のアプローチ

一方、Rector[Rector 02]は、医療の巨大なオントロジー構築経験から、対象の内容を峻別するのは人間であるが、そうして表現されたオントロジーが大規模になったとき、メンテナンス性が落ちることを考慮するという立場から、記述論理を背景にして、表現された内容を整理して機械が扱いやすくする基盤表現を提供している。

(3) 我々のアプローチ

それらに対して、我々は、最終的に概念を峻別するのは人間ではあるが、概念の峻別は、高度なオントロジーリテラシーが要求されるため手助けすることが必要であると考えている。そのた

めには、人間が理解しやすい表現と、機械が扱いやすく概念化した表現の両方を扱えるようにする必要があると考え、人間が認識しやすいグラフィカル表現と機械が扱いやすい形式表現を対応付けたオブジェクトとして提供することを試みている。

3.2 概念峻別の手掛かり

人間が概念を認識するとき、単一の概念だけではなく、共通点を持つ概念を集合として捉えることがある。例えば、成人と言えば、同時に、子供や老人も同時に想起する。それらは「成長段階」といったまとまりとして括られる概念の集合として認識されている。さらにそうした概念の集合で捉えると、概念間には何らかの制約があることが分かる。先ほどの例では、成人が、同時に子供や老人になることはないし、成人から子供に戻ったりすることはないなどが認識できる。

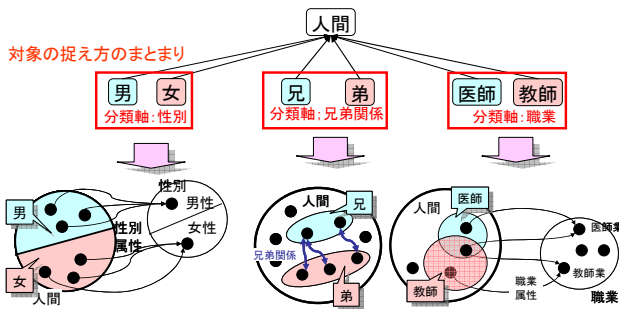


図 3 概念の分類軸とその集合表現の違い

そこで我々は、図 3 のような集合を考え、こうした何らかの共通点を持つ概念の集合をひとつの単位として捉えたとき、何を共通点としているのか？その共通点が認識されたとき、その所属概念はどんな性質を持つのか、概念間にはどんな制約があるのかなどを規定できると仮説設定し、[上田 07]では、is-a階層において、その分類軸を基点にして、大別して 4 種類の認識パターン（現在は 5 種類）があることを示し、認識パターン毎に OWL において各概念やプロパティの制約を表現できることを明らかにしている。

3.3 概念化アスペクト

上記の認識パターンの分析から得られた知見に基づき、人間が概念の共通の性質をまとめて捉えようとする対象の捉え方に注目し、人間がどのように対象を捉えているかを表現しやすいフレーム表現と機械によって記号処理可能な表現を対応づけて概念化したものが概念化アスペクト(CA)である。図 4 に CA

の例として性別区分CAを示す。性別区分CAは生物の性別という属性に注目したとき、その属性値によって男性、女性と認識され、このとき生物は、男性、女性のどちらでもないこと、男性かつ女性とは認識されないこと、性別は変わらないことが概念化されている。こうしたことをCAのインスタンスであるCAオブジェクトで表現される。

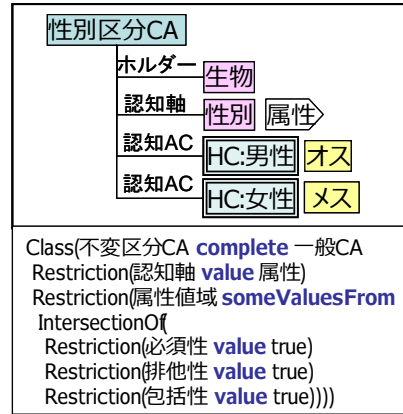


図 4 概念化アスペクト(CA)オブジェクトの例

CA オブジェクトは、人間向けの表現であるフレーム表現と、その構成要素の制約が機械可読な形で概念化した部分から構成される。

まずフレーム表現は、CA タイプにもよるが、おおよその基本構造は以下の通りである。CA オブジェクトが表す概念の捉え方が明示される。まずホルダーは、CA を適用可能なクラス制約を表す。次に認知軸は何に注目して CA が概念化されているかを表し、認知軸の概念型とともに指定される。最後に認知 AC は、ホルダーが CA の基で認知されたときの低位概念の性質を表し、性質をアスペクト概念として記述される。

一方の機械可読表現では、そのCAオブジェクトの構成要素が満たす制約が記述される。実際は、CAタイプにタイプ付けされており、CAタイプ側で定義されている。図 4 では、固定区分CAの注目する属性の属性値は必須であり、属性値間は排他で包括的であることが示されている。これらはOWLで表現する際のプロパティ制約、クラス制約として利用される。

3.4 概念化アスペクトの意義

CA を概念化したことによって、対象のとらえ方とその表現をモノとして扱えるようになる。これはオントロジーエキスパートが

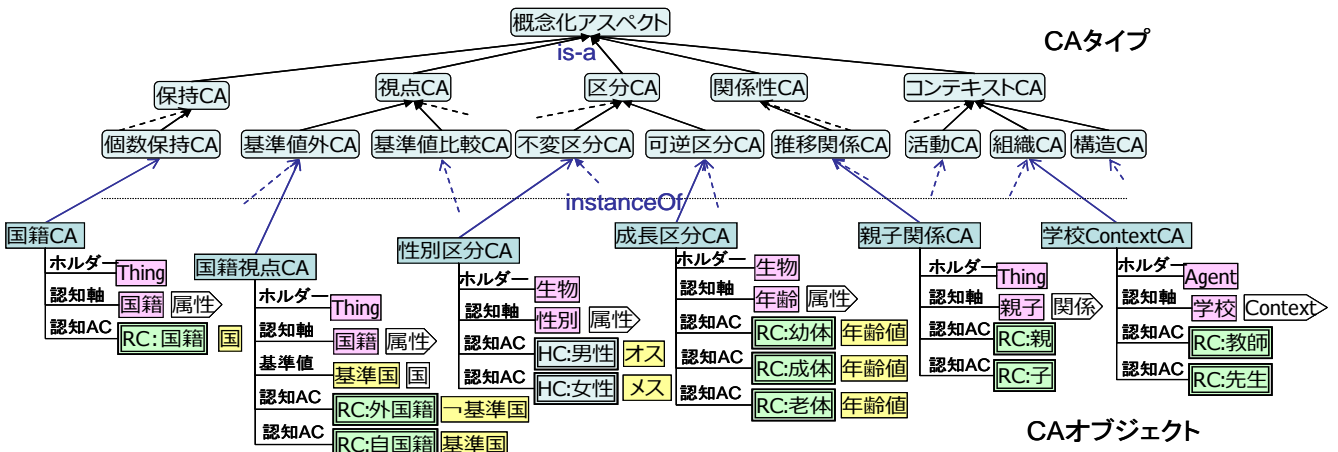


図 5 概念化アスペクトの体系例

持つ対象のとらえ方と表現を CA オブジェクトとして記述すれば、オントロジー構築ノウハウとして蓄積・共有できることを意味する。こうした CA 共有システムを組織内で運用すれば、組織としてのオントロジー運用力を高めることが期待できる。図 5 にそうして蓄積される CA の体系例を示す。CA タイプは、共通性の分析に基づいて定義される概念定義であり、現在五つの大分類を用意し、その細分類を定義している。CA オブジェクトは、CA タイプをインスタンス化し、具体的な共通性とそのメンバー概念などを記述される。こうした CA の体系がタスクや領域毎に構築されることが期待される。

4. 概念化アスペクトに基づくオントロジー構築支援

我々が想定するオントロジー構築支援は、ツールを提供することによって単に構築支援を行うだけではなく、組織内でオントロジー構築・利用に関する知識共有サイクルを回せるようにすることにポイントがある。このとき中心となるのが概念化アスペクトであり、組織内のオントロジーエキスパート(OE)のノウハウを CA オブジェクトとして蓄積・共有することで、組織内のオントロジーリテラシーを向上させ、構築・利用能力を高められるのではないかと考えている。

以下、そのようなオントロジー構築利用サイクルの観点から我々が想定している CA の 3 種類のオントロジー構築支援への適用案を説明する。

4.1 CAオブジェクトの抽出(図 6(a))

既存オントロジーから共通認識にあたる部分から CA オブジェクトを抽出する。システムはオーサが共通性を持つと認識した概念集合がどの CA タイプに該当するのか判定を CA の機械可読表現を用いて支援する。この作業は、オントロジーリテラシーを要求するため、オントロジーエキスパート(OE)が行うことを想定している。抽出された CA オブジェクトは、CA リポジトリに蓄積され、他のオントロジーオーサがオントロジー構築のノウハウとして利用できるようになる。

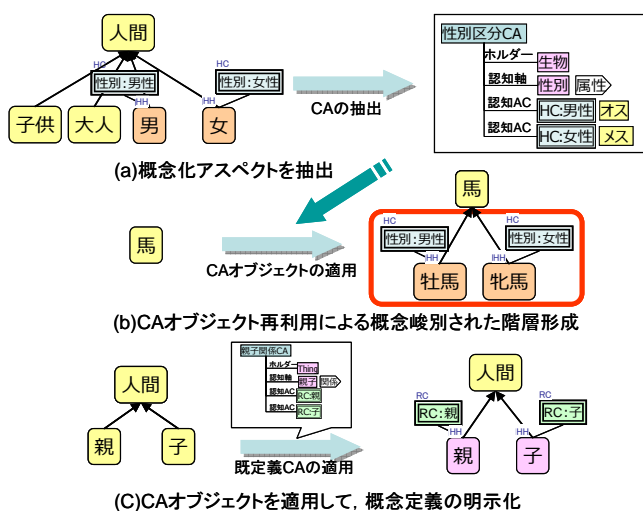


図 6 概念化アスペクトを用いた概念定義支援

4.2 CAを用いた概念階層生成(図 6(b))

オーサは、基本概念(BC)に対して CA オブジェクトを当てはめることで、概念階層を生成できる。このときシステムによって基本概念(BC)は CA オブジェクトのホルダー制約を満たしている

かが確認され、認知 AC を保持するアスペクトホルダー(AH)が BC の下位概念として生成される。オーサは生成された AH にラベルをつければよい。これにより下位概念の概念タイプが明示され、オントロジーの正規化プロセスで、概念の峻別が自動的にできるようになる。

4.3 CAによる概念定義の明示化(図 6(c))

(b)と基本的に同一であるが、既定義の概念階層に対してCA オブジェクトを適用することによって概念の峻別を行う。概念の峻別がなされれば、2.3で示したオントロジーの正規化を使って、既定義オントロジーの正規化を行うことができるようになる。

5. 最後に

本稿では、オントロジーリテラシーを持たないオーサによるオントロジー構築を支援するために、人間が理解しやすい表現と記号処理可能な表現を対応づけて概念化した概念化アスペクト(CA)に基づくオントロジー構築支援を提案した。CA によってオントロジーエキスパート(OE)のもつ対象の捉え方に関するオントロジー構築ノウハウを人間および機械可読な CA オブジェクトとして表現でき、それらを共有することで OE 以外のオーサによる概念の峻別を支援できるようになる。概念の峻別ができれば、概念表現の制約から、オントロジーの再利用性を考慮した形に正規化ができるようになることを示した。

また CA オブジェクトを適用したオントロジーは、その構築意図が明示されることで、オントロジーの可読性がある。それによりオントロジーの合意形成やメンテナンスにおける手掛かりが得られることになる。

現在、オントロジー正規化についてはプロトタイプを作成し、その理論の検証を進めており、Swoogle[Swoogle]のようなOWL オントロジーリポジトリにある既存オントロジーを対象に正規化の検証を行う予定である。さらに今回提案したCAによる概念峻別の支援を含めた理論を実証するために、CAを編集・利用できる環境の構築を進めることを予定している。

参考文献

[溝口 05] 溝口理一郎:オントロジー工学, オーム社, 2005.
 [溝口 06] 溝口理一郎:オントロジー構築入門, オーム社, 2006.
 [Rector 02] Rector, A. : "Normalisation of ontology implementations: Towards modularity, reuse, maintainability", In EKAW Workshop on Ontologies for Multiagent Systems, 2002.
 [Swoogle] Swoogle Semantic Web Search Engine, available at <http://swoogle.umbc.edu/>
 [上田 06] 上田, 池田: 概念化アスペクト:オントロジー構築の手がかり, 第 14 回SWO研究会, SIG-SWO-A602-03, 2006.
 [上田 07] 上田, 池田: 概念化アスペクトに基づくオントロジー構築プロセスのための概念表現モデル, 信学技報 Vol.106, No.472, KBSE2006-50, pp.31-36, (2007).