

# 「法造」におけるオントロジーの品質保証に関する一考察

## A Consideration of Quality Assurances for Ontology in “Hozo”

太田 衛<sup>1,2</sup> 古崎 晃司<sup>1</sup> 溝口 理一郎<sup>1</sup>

Mamoru Ohta<sup>1,2</sup>, Kouji Kozaki<sup>1</sup>, and Riichiro Mizoguchi<sup>1</sup>

<sup>1</sup>大阪大学 産業科学研究所

<sup>1</sup>The Institute of Scientific and Industrial Research (ISIR), Osaka University

<sup>2</sup>株式会社 エネゲート

<sup>2</sup>Enegate Co., Ltd.

**Abstract:** Recently, ontologies are being constructed in various technical domains. The quality of an ontology is one of the important factors that determines its utility. In order to assure its quality, not only form-based evaluation as to whether the ontology being constructed is written properly in terms of its form (syntax), but also content-based evaluation as to whether the ontology properly represents the target domain, whether the ontology actually serves for problem solving, etc. is necessary. In this study, we consider the quality assurance of ontologies in Hozo, which is an environment for building/using ontologies that is being developed by the authors. As to form-based evaluation, Hozo provides various assistance functions for properly editing an ontology in compliance with several guidelines and rules. As to content-based evaluation, Hozo introduces a method for supporting ontology evaluation through conceptual maps which are generated from the ontology according to the user’s viewpoints.

## 1. はじめに

近年、専門領域の情報化が進み、専門知識が細分化される中で、様々な専門領域で知識の構造化へのニーズが顕在化しつつある。知識の構造化を行う手段のひとつとしてオントロジー工学が注目されており、様々な専門領域でオントロジーの構築が進められている。このような背景のもと、オントロジーの専門家のみならず、各ドメインの専門家が適切にオントロジーを構築できるよう支援するための方法論やツールの開発が求められている。

オントロジー構築のプロセスは設計、構築、洗練の3つのフェーズに大きく分かれる。構築するオントロジーの品質は、その機能を決定づける重要な要素のひとつである。オントロジーの品質を保証するには、オントロジー構築の各フェーズでオントロジーが正しく書けているかどうかを評価し、その結果を構築作業に反映させる必要がある。一般的にオントロジーの評価は、形式的な評価と内容的な評価の2つの評価に大きく分けられる。形式的な評価に関するアプローチとしては、OWLにおける記述論理(DL)による推論処理を用いた矛盾検出などの機械的なアプローチをとることが一般的である。一方、内容的(意味的)な評価に関しては、構築したオン

トロジーに基づいたモデルを実際の対象世界の問題解決に適用して有効性を示す方法や、ドメインの専門家による内容評価を行う組織を形成する取り組み[1]など人的なアプローチを取る方法がある。

本論文では、筆者らが開発を進めているオントロジー構築・利用環境「法造」<sup>1</sup>における、オントロジーの品質を保証する為の枠組みについて述べる。2章ではオントロジー構築プロセスにおける品質保証の位置づけについて述べる。3章では形式的な評価のための編集支援と整合性検証の枠組みについて述べる。4章では内容的な評価のための概念マップを用いた洗練支援手法とその実践例について述べる。5章では関連研究との比較を述べる。最後に6章では本研究の成果をまとめ、今後の検討課題を述べて総括する。

## 2. オントロジー構築プロセスにおける品質保証

オントロジー構築のライフサイクルは、大きく設計、構築、洗練の3つのフェーズに分かれる。本節では、オントロジー構築の各フェーズにおいて想定

<sup>1</sup> <http://www.hozo.jp/>

される品質保証の取り組みと、それを支える枠組みについて述べる。本研究ではオントロジーの品質保証のうち、オントロジー構築・洗練過程における品質保証を対象とする。

## 2.1 オントロジーの設計

オントロジー設計フェーズでは、ドメインの専門家とオントロジーの専門家が協力して、対象領域における主要概念（知識）を抽出、整理し、上位概念構造の検討など、オントロジーの基本設計を行う。その過程では概念定義を行う上で、何が本質的かといった本質属性や、概念の共通性・汎用性などを適切に捉えるためのオントロジー工学的な考察が重要となる。設計時における品質保証の取り組みとしては、既存のオントロジーにおける設計事例をデザインパターンとして蓄積しておき、設計中のオントロジーと照らし合わせることで、設計原理や事例の再利用を検討するという手法がある。Presuttiらはオントロジー設計の実践に使用できる基本構成要素としてあらかじめ登録されているデザインパターンから、推奨されるものを提示して設計時に再利用することを支援するツールを提供している[2]。

また、上位概念構造の検討には、特にオントロジー構築の経験的や哲学的な考察などが必要とされ、ツールで直接的に支援するのは困難と思われる。そこで、上位概念構造として既存の上位オントロジーを採用して設計に活かすケースもみられる。実際にDOLCE<sup>2</sup>、BFO<sup>3</sup>、GFO<sup>4</sup>、SUMO<sup>5</sup>、CYC<sup>6</sup>、YAMATO<sup>7</sup>といった上位オントロジーが、様々なオントロジー構築で利用されている[3]。

## 2.2 オントロジーの構築

上位概念構造の設計がある程度固まってきたら、上位概念の定義に従って、各概念を増やしていき、ミドルレイヤー以降の概念構造を組織化する。構築時における品質保証の取り組みとして、構築したオントロジーが形式的（文法的）に正しく書いているかを検証する必要がある。その際には、個々の概念定義が記述形式に基づき正しく書いているかとともに、オントロジー全体として各概念定義間に不整合性（矛盾）がないかの検証が必要とされる。

このようなオントロジーの形式的な評価

(verification) は、記述言語の仕様に対応した推論機構などを用いることで、その規約をチェックするような機械的処理が実現できる。一般的に構築者が記述形式を正しく理解し、その整合性を検証しながらオントロジーを構築するには作業負担が大きい。よって構築ツールでは、構築者の作業負担を減らし規約どおり編集するための支援機能の実現が重要とされる。

## 2.3 オントロジーの洗練

オントロジー洗練過程の中では、構築したオントロジーが対象世界を適切にモデリングできているかを、概念定義の内容を確認して評価し、問題点を修正する。洗練時における品質保証の取り組みとしては、2.2節で述べた形式的な評価に加え、オントロジーが内容的（意味的）に正しく書いているかどうかを評価する必要がある。その際には、ドメインの専門家による評価、ユースケースに基づいた検証、実際の問題解決に適用した結果による評価といった手法が採られる。内容的な評価(validation)に関して、現実世界のモデルと照らし合わせた上で専門家の知識・経験を通しての判断が必要な為、人手による洗練作業を中心に行っているケースが多くみられる。

ドメインの専門家が内容の評価を行うには、第一にオントロジーの定義内容を適切に理解する必要がある。そのために専門家の視点でわかりやすくオントロジーを提示する必要があり、オントロジーの可視化やドキュメント化によって人間が直感的に理解しやすくする手法が用いられることが多い。

ユースケースに基づいた評価手法の一つとして、OntoEditではOn-To-Knowledgeという方法論に基づき設計時に記述された能力質問(Competency Question)に対してオントロジーが答えられるかどうかを確認するための枠組みを提供している[4]。また実際の問題解決に適用した上で評価するには、オントロジーに基づくシステムを開発し、そのシステム性能を通して評価される。その為には、オントロジーを用いたシステム開発を効率的に行うAPIなどのソフトウェア環境が必要になる。

## 3. 法造における形式的評価の枠組み

### 3.1 オントロジーの形式的評価の概要

2.2節で述べたとおり、オントロジー構築時における品質保証のために、構築ツールには主にオントロジーの規約通り正しく編集するための支援機能が求められる。ここで計算機によるオントロジーの形式的な評価と修正支援を行うには、編集時に随時エラーをチェックしてエラーを提示する方法と、編集後

<sup>2</sup> DOLCE: <http://www.loa-cnr.it/DOLCE.html>

<sup>3</sup> BFO: <http://www.ifomis.org/bfo/>

<sup>4</sup> GFO: <http://www.onto-med.de/ontologies/gfo/>

<sup>5</sup> SUMO: <http://www.ontologyportal.org/>

<sup>6</sup> CYC: <http://www.opencyc.org/>

<sup>7</sup> YAMATO: [http://www.ei.sanken.osaka-u.ac.jp/hozo/onto\\_library/upperOnto.htm](http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm)

に整合性を検証してエラーを修正する方法の2つアプローチがある。構築ツールではチェックする内容に応じて、編集時/編集後のいずれかのタイミングでオントロジーの整合性をチェックするのが適切かを考慮した上で、支援機能を実現する必要がある。

法造ではこれら2つのアプローチを適宜使い分け、編集時には必要最低限のエラーを回避する為の編集支援を重点に置き、編集後にオントロジー全体の整合性検証を行う枠組みを採っている。その理由は、編集時に厳密な整合性チェックを行うと、構築の途中段階でエラーが頻発して、ツールの操作性が下がりユーザの発想を阻害する恐れがあるためである。

そこで形式的な評価の枠組みを法造に導入するにあたって、まずはオントロジーの規約に対して検証項目を数え上げた(付録A, 全33項目)。それらをオントロジー編集時の支援機能として利用者の発想を阻害しない程度で、エラーを提示することが好ましい最低限守るべき項目(付録Aの表の[編集時]欄)と、1つの編集操作ではエラー判断がしにくいなどの理由で、編集後にオントロジー全体の整合性を検証してエラーを提示すべき項目(付録Aの表の[編集後]欄)とを分けた。

これらの整合性チェックを保証する際に、法造では完全性よりも高速性を重視した目的指向の推論処理を採用している。具体的には、付録Aで示した33の検証項目に対して個別に整合性チェックする手続き的な推論処理で実現している。個々の推論処理は単純で説明不要であろう。一見、全てのオントロジーの検証項目を編集後に整合性チェックしているように見えるが、表で示したとおり、いくつかの検証項目は編集時にもチェックをしている。

### 3.2 オントロジー編集支援

法造ではProtégé<sup>8</sup>などの多くのツールで採用されているis-a階層とGUIを用いた編集機能のみならず、ユーザが記述内容を直感的に捉えることができる独自のグラフィック表示による編集を採用している。さらに、オントロジー構築者が規約どおり正しく編集できるように各種の支援機能を提供している。

例えばスロット編集時には、上位概念からスロットの定義内容を継承して定義内容を変更する操作(スロットの特殊化)やスロットのクラス制約や値を編集する際に、専用ダイアログでのガイダンス等を用いて概念間の整合性を保ちつつ正しく定義できるように設計している。例えばスロットの特殊化支援では、規約通りに正しく記述する為にスロット編

集時に上位概念から継承可能なスロットの候補をリストで提示し、選択編集できるようにしている。その一方で未定義概念の参照などの特定の項目については正しい記述を強要するのではなく、表示色を変えて警告のみを提示するなど、利用者の発想を阻害しないよう配慮している。

また概念名の変更や継承関係の変更といった、記述内容の修正時に、概念間の整合性が崩れるのを防ぐ支援機能も実装している。例えば概念名変更時の追従変更支援では、概念名を変更した際に、その概念を参照している全ての箇所においても概念名を同時に追従して変更し、オントロジーとしての整合性が保たれるようにしている。

このように法造では新規の概念を記述する時だけでなく、既に記述されている概念やスロットを修正する際にも、オントロジーとして矛盾がないよう整合性を保つ為の支援機能を用いて、その品質を保証するようにしている。

### 3.3 オントロジー編集後の整合性検証

3.2節で述べた編集支援により、ある程度のエラーを避けることができるが、1つの概念を編集している際には整合性が判断できず避けることが出来ないエラーがある。更に概念定義の変更や削除を繰り返すうちに、その概念間の関係性が変更され、オントロジー全体の整合性が損なわれる場合もある。また編集時に警告のみで許容したエラーの修正漏れが残っている場合もある。これらを解消する為には編集後にオントロジー全体の一貫性を保証する仕組みが必要となる。

そこで編集時の作業支援でカバーしきれない検証項目に対して、オントロジーの編集後に全体の整合性を検証し、その結果を提示するオントロジー整合性検証機能を実装した。さらに、検証結果として提示されたエラーを取り除く為の修正操作を専用ウィザードにより支援する修正支援機能を開発した。なお、これらの整合性検証機能は、法造独自の理論的な枠組みを計算機で処理するために開発したソフトウェアモジュールである、「オントロジー利用API:法造コア」と「推論機構」を利用して開発している<sup>9</sup>。

### 3.4 オントロジー編集支援の評価

本節では、3.2節で述べた形式的評価として導入したオントロジー編集支援がオントロジー品質向上にどれだけ貢献したかに関する評価実験について述べる。

<sup>8</sup> <http://protege.stanford.edu/>

<sup>9</sup> 両モジュールは、フリーソフトウェアとして法造のサポートサイト(<http://ontsupport.enegate.jp/ontology/>)において公開されている

### 3.4.1 評価方法

オントロジーの品質向上に対する評価実験として、3.2 節の編集支援機能の改良前後でオントロジー構築作業の形式的な観点での品質向上が図れたかどうかを比較検証した。対象者は初心者としてオントロジーに関する知識を全く持たない大学院生を対象とした。実験手順として、まず最初に、オントロジーについての説明及び法造の使い方に関する講習を事前に行った。次に、ある一つのテーマ（「乗り物オントロジー」）に対して、知識の個人的なバラツキがないように3名を1グループとしてツールを使ってオントロジー構築をさせた。形式的な観点での品質向上を検証するためにツールの改良前後でそれぞれ同じ内容で評価実験を行った。実験結果として構築されたオントロジーの基本概念数、スロット数、関係リンク数、形式的なエラーの数を集計した。また、形式的なエラーの数に関して3.3 節で述べたオントロジー整合性検証機能を使用した。

### 3.4.2 結果と考察

実験結果は表1のとおりである。被験者数が異なる為、実験結果のすべての値は全グループの平均をとって算出している。改良後にオントロジー編集支援機能を使用して構築されたオントロジーは改良前に比べて、基本概念数、スロットの生成数、関係リンク数はともに増加した。またスロットの特殊化の数は著しく増加した。これは編集しにくかったスロットの特殊化操作の改善による結果であり、概念の継承関係とそれに伴い、必要な意味のあるスロット定義を数多くできたとみることできる。しかも構築された概念や関係が増えたにも関わらず、オントロジーの形式的なエラーは減少している。以上の実験結果は形式的なエラーが減って、質の高い概念や継承関係を初心者である被験者が数多く定義できたことを示しており、3.2 節で述べたオントロジー編集支援機能がオントロジーの品質向上に寄与しているといえる。

なお、法造でのオントロジー構築経験のあるユー

表1 オントロジー編集支援の評価

対象	基本概念の生成数	スロットの生成数	スロットの特殊化数	関係リンクの生成数	エラー率
改良前 (※1)	27.27	21.09	9.82	0.82	10.30%
改良後 (※2)	<b>34.38</b>	<b>29.54</b>	<b>17.54</b>	1.23	4.94%
増減(%)	<b>26.08%</b>	<b>40.05%</b>	<b>78.63%</b>	50.43%	<b>-52.09%</b>

※1:機能改良前の評価実験(13グループ:39名)

※2:機能改良後の評価実験(15グループ:45名)

ザに関しても「以前よりも使いやすくなった」と操作性の向上に関して肯定的な意見を得ている。

## 4. 法造における内容的評価の手法

2.3 節で述べたとおり、オントロジー洗練時にドメインの専門家が内容評価を行うには、まずオントロジーの定義内容を適切に理解する必要がある。オントロジーの理解促進のための枠組みがあればドメインの専門家の理解や協力も得やすくなり、良質なオントロジー構築にも貢献すると考えられる。オントロジーの理解促進に関してオントロジーの俯瞰(利用)を目的とした先行研究では、専門家の関心に応じて様々な視点からマップを生成する基本的な枠組みを提案している[5]。そこで本研究ではオントロジー洗練時の理解支援とした手法として概念マップを用いることとする。本研究が目的とするオントロジーの構築・洗練においては、その注目する視点がオントロジーや評価目的によって変わり、概念マップを生成する際に視点を絞り込む必要がある(図1)。すなわち、概念マップによるオントロジーの構築・洗練の検討課題としては、評価目的に応じた概念マップ生成のために、どのような視点を選べば良いか、また抽出した概念間の関係性をどのように可視化すれば効果があるのか等の検討が必要とされる。

次節では、本手法の適用事例を通して、これらの検討課題と本手法の有効性について考察する。

### 4.1 サステナビリティ学オントロジー構築における適用

RISS(大阪大学サステナビリティ・サイエンス研究機構<sup>10</sup>)における知の構造化研究グループでは、環境分野におけるサステナビリティ学に関係する多様な領域の知識を、領域に依存しない形式で体系化したサステナビリティ学に関するオントロジーの構築に取り組んでいる[6]。その構築過程においては、構築したオントロジーから概念マップを生成し、対象とする知識が適切なマップとして生成できるかを確認することでオントロジーの構築・洗練を行ってきた。また同様な手法を、環境省「Hc-082 アジア太平洋地域を中心とする持続可能な発展のためのバイオ燃料利用戦略に関する研究」という別プロジェクトにおいてバイオ燃料分野に関連する概念を中心に既存のサステナビリティ学オントロジーを拡充する際にも適用した。その際には環境分野の専門家が作成した29の典型シナリオをもとにオントロジーを拡充し、それらのシナリオが概念マップとして再現出来ているかを確認することでオントロジーの

<sup>10</sup> <http://www.riss.osaka-u.ac.jp/>

洗練を行った。マップの再現には先行研究[5]で開発したオントロジー探索ツールを使用した。シナリオをもとに拡充したオントロジーの概念数は 649 から 1,892 に、スロット数は 1,075 から 2,119 に増加した。また構築に使用した 29 の典型シナリオのうち 27 がマップとして再現して確認でき、その再現率は 93% となった<sup>11</sup>。

具体的な作業としては、まず典型シナリオからシナリオに現れる概念及び概念間の関係性を抽出し、新たに概念やスロットを追加するという拡充作業を行った。例えば、シナリオ「燃料作物農園開発のための森林伐採の際の火入れやサトウキビ収穫時の火入れによる大気汚染、森林火災の原因、土壌劣化、水質汚濁」からは、下線部の「燃料作物農園開発」や「森林伐採」といった新たな概念を抽出しオントロジーに追加した(図2)。次に、その概念追加したオントロジーを対象にシナリオの流れに沿って探索を行い、概念マップを生成した。この概念マップにより、シナリオで想定した概念間の意味的な繋がり(パ

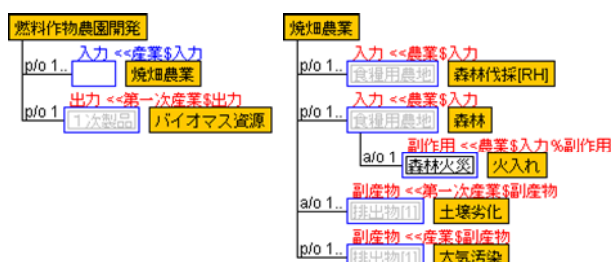


図2 シナリオに基づくオントロジー拡充作業例

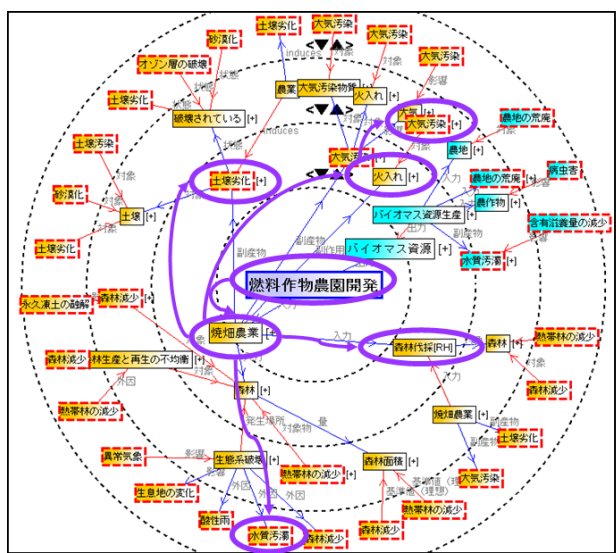


図3 オントロジー洗練における概念マップの活用事例

<sup>11</sup> 確認できなかったシナリオに関しては、探索上の不備ではなく、オントロジー構築上の不備があった為、マップとして再現できなかった。

ス)を確認して、オントロジーの内容的な評価を行った。先の例で示したシナリオを概念マップで再現したものを図3に示す。概念マップ上の太線で囲ったノードが拡充した概念(スロット)である。この構築・洗練作業では、概念マップとして表したいシナリオが多様な内容であることから、概念マップ生成に用いる視点には特に制限を設けず、利用者が自由に視点を選択してマップを生成するという手法を用いた。

このように既存のオントロジーとの整合も取りながら拡充したオントロジーの意味的な評価を行うことができ、この構築・洗練作業には有効であることが分かった。

## 4.2 臨床医学オントロジー構築における適用

厚生労働省の「医療情報システムのための医療知識基盤データベース研究開発事業<sup>12</sup>」では、臨床医学の複数領域にわたる電子データの統合管理や高度な情報解析を行う情報基盤技術として不可欠であるとして、国内初の本格的な臨床医学オントロジーの構築が進められている[7, 8]。臨床医学オントロジーの構築では、初期フェーズとして臨床医学の専門家と共同作業で臨床医学に必要とされる主要な概念が記述できるかどうかの検討を行い、具体的には疾患および疾患に関連する症状・所見、および人体構造知識を中心として、その基底となる上位概念構造の基礎的な考察を行っている[9]。また中期フェーズとして、これら共通の上位概念構造をもとに、疾患および人体構造に関するオントロジーを専門家が構築している。

これらの構築・洗練作業では、ドメインの専門家がオントロジーの内容を確認する為、共通の上位概念構造のうち注目した視点のみに限定して可視化した概念マップを生成するという手法が求められた。次節以降では臨床医学オントロジー構築における概念マップ生成の詳細について述べる。

### 4.2.1 着目した上位概念構造の関係性に基づく可視化

臨床医学オントロジーの構築・洗練作業では、2.2節で述べた形式的な評価とともに内容的な評価を行う必要があった。そこでオントロジー構築・洗練支援のための概念マップの生成方法として、臨床医学オントロジーの共通の概念構造を持つ概念群に対して、視点として着目した概念間の関係性だけを抽出

<sup>12</sup> <http://www.m.u-tokyo.ac.jp/medinfo/medont2009proj/>

して可視化する手法を採用した。これにより、オントロジーの構築・洗練時にドメインの専門家が評価したい内容だけを可視化し確認しやすくなる(図4)。

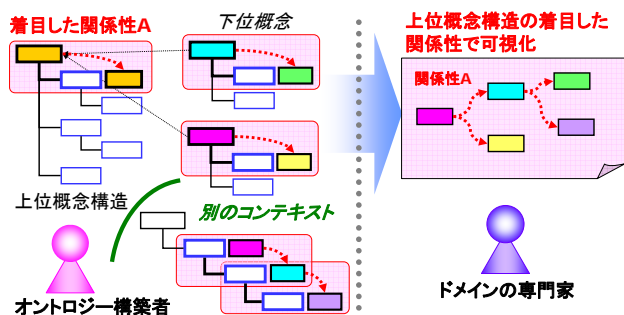


図4 着目した上位概念構造の関係性に特化した可視化

#### 4.2.2 人体構造の接続オントロジーの構築・洗練支援

##### (1) 接続関係の定義

臨床医学オントロジーにおける人体構造の接続関係は、「ポート」という共通の上位概念構造を導入し、血管や神経といった流体や信号を通すような機能的な接続、骨と関節の接続といった力学的な接続、それ以外の位置関係を表すような空間的な接続<sup>13</sup>など、様々な接続関係を同型のフレームで扱えるように工夫されている。これにより血管系や神経系といった機能的な接続関係の繋がりや骨どうしの繋がりを計算機で推論することが可能になっている。例えばある血管Aと血管Bの接続関係を、人体構造物とポートの概念を用いて表現した場合、接続ポートの接続先としてお互い接続ポートを相互参照する形で記述される(図5)。接続の種類は、接続の記述に用いるポートの種類の違いで表している。さらに関節や血管の分岐点など、複雑な接続構造については関与する人体構造を仮想的な部品に分割するなどの手法を用いて、最小単位の接続に分解することにより、ポートを用いた同型の接続構造の組み合わせで捉えている。

現状の臨床医学オントロジーでは人体構造の接続関係について、心臓から器官までの血液の流れを対象とした循環器系統(動脈系、静脈系)、全身の主要な神経系統、骨格筋の接続関係のオントロジーが構築されている。各接続関係のオントロジーで定義されている概念数は、動脈系、静脈系を合わせた循環器系統で約11,000、神経系統で約7,500、またそれらの間の接続関係の数は、それぞれ約8,600、約3,200

<sup>13</sup> 位置関係や空間的に近接しているといった本来は接続とは言えない概念も、共通して枠組みで捉えるためにポートを用いた同形の概念構造で捉えている。

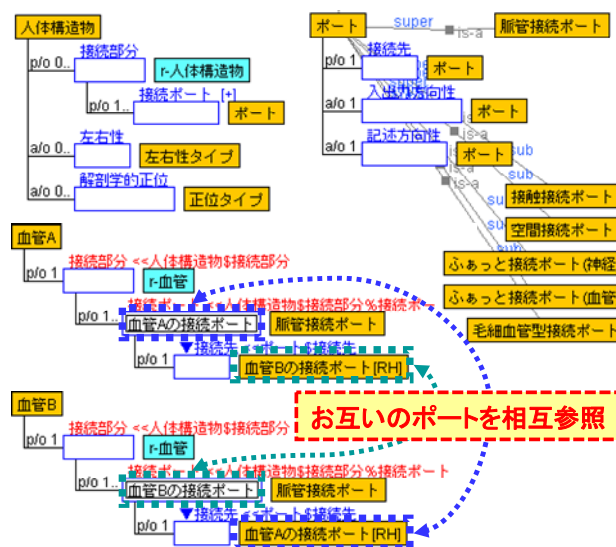


図5 人体構造における接続関係の記述

となっている。

##### (2) 人体構造の接続オントロジーの構築・洗練支援

前項で述べた人体構造の接続オントロジーにおいて、接続関係が内容的に正しく記述されているかどうかを確認するには、人体構造物の接続ポートの接続先に記述されている内容を順番に辿って確認すればよい。例えば、図5で示した血管Aと血管Bが接続されているかを確認するには、{血管A→Aの接続部分→Aの接続ポート→Bの接続ポート→Bの接続部分→血管B}と辿ればよい。すなわち、循環器系統において心臓から送り出された血液が動脈の血管を通して全身の器官や組織に供給され、静脈の血管から心臓に戻ることが正しく表されていることを確認するには、このようなポートに沿って接続関係を辿る作業を接続の数だけ繰り返す必要がある。

しかし現実的な問題として、1万を越える人体構造の接続関係を概念定義の内容を手動で辿って確認するのは事実上不可能に近い。よって、このような人体構造の接続関係が正しく記述されているかを確認し、オントロジーを洗練するには、必要な探索を自動的に行い、利用者(専門家)が把握しやすい形式で可視化することが求められる。しかもこのような支援は共通性のある概念定義の場合には機械的な推論処理での対応が可能な範囲であり、その作業負担の軽減や利用者の内容理解に対する貢献度も高い。

##### (3) 人体構造の接続関係の可視化ツールの試作

そこで、このような必要性に応えるため、人体構造を記述したオントロジーから接続関係を表す「接続ポート」を辿るような探索を行い、その探索結果を人体構造の接続関係の確認用に可視化する専用ツールを試作した(図6)。この可視化ツールの基本操



図 6 人体構造の接続関係の可視化ツールの画面例

作として起点となる人体構造物を選択し、そこから画面右下のリストに表示される接続先の人体構造物を選択することで、接続関係を辿って逐次表示していく。例えば、図 6 の画面例では、「心臓」を起点とする「大動脈」の接続関係を可視化しており、「心臓」から「脈管接続」を辿ると、「大動脈分岐」で分岐して「大腿動脈」まで「動脈」が繋がっていることが可視化によって理解できる。また「右外腸骨動脈」からは「ソフト接続<sup>14</sup>」を介して「右大腰筋」といった筋組織へと血液が供給されていることもわかる。

このように臨床医学オントロジーの人体構造物の接続関係の可視化を行うことによって、記述内容が直感的に理解しやすくなり、概念記述のエラーも見つけやすくなった。実際に専門家と共に「動脈」の接続関係を、可視化ツールを使って確認したところ、本来繋がっているべき血管の接続関係が記述できていないことが発見できた。

なお、可視化のための関係性の探索は、4.2.1 項で述べた考え方に基づいて人体構造における接続関係の共通の上位概念構造を用いて視点に基づき、関係の種類によらず統一した処理で実現されている。また、この接続関係の可視化ツールはJavaのクライアントアプリケーションとして開発し、概念間の関係ネットワークの可視化には、オープンソースのグラフ描画ツールJUNG<sup>15</sup>を利用しているが、オントロジー構築・洗練支援のためにより最適な可視化方法については検討を継続している。

### 4.2.3 疾患オントロジーへの適用

4.2.2 項で試作した可視化ツールを臨床医学オントロジーにおける疾患概念（疾患オントロジー）の構

<sup>14</sup> 血管から毛細血管を解して各組織へ血液が流れていることを表す接続。現時点での名称は仮称であり、図中では「ふぁつと接続」となっている。

<sup>15</sup> <http://jung.sourceforge.net/>

築・洗練にも適用して検証した。

### (1) 疾患概念の定義

臨床医学オントロジーにおける疾患は、「その原因と途中経過を含めた一連の状態変化の連鎖と、それにより引き起こされている結果状態との総体」として捉えられ、疾患概念はその疾患を発症した患者には共通して見られる状態変化（連鎖）と定義される[8]<sup>16</sup>。よって疾患概念の定義では、その疾患で注目している異常状態の状態変化のみが記述される。一方、疾患を構成する「異常状態」は患者に起こりうる状態として一般化し、その「原因」や「結果」となる異常状態を属性として持つ疾患概念とは独立した概念として定義される（図 7 (a)）。これにより、ある患者に起こり得る異常状態の状態変化は、各異常状態の原因・結果を辿ることによって得ることができる（図 7 (b)）。

オントロジー構築時には、起こり得るすべての異常状態の変化を予め数え上げて定義しておくことは、一般的に不可能と思われる。そこで、疾患定義の際には、医師がその疾患で注目する異常状態とその状態変化だけを記述した。そこから記述された異常状態全てを自動抽出し、それらを一般化する事によって自動的に一般化された異常状態を収集するという手法をとった。これを多くの疾患について行うことにより、ある疾患では注目されなかった異常状態の連鎖が、別の疾患では注目されることがあるので、結果的に対象とした全疾患の範囲で考えられ得るすべての異常状態の連鎖が汎用的な形で自動的に定義される。この汎用的な連鎖を用いることで、ある疾患定義で注目され記述された異常状態の連鎖以外にその患者に起こる可能性のある一連の状態変化（連鎖）の総和を得ることができる。例えば、疾患「糖尿病」に定義において注目される異常状態としては「高血糖状態」しか定義されていなかったとしても、そこから派生する一連の状態変化は異常状態「高血糖状態」の結果状態「網膜血管障害状態」、更にその

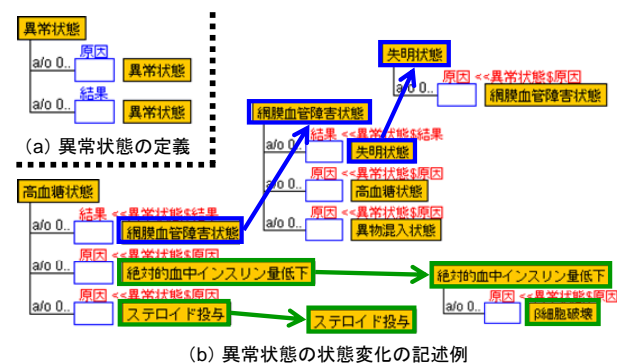


図 7 疾患における異常状態の連鎖の記述例

<sup>16</sup> <http://www.m.u-tokyo.ac.jp/medinfo/medont2009proj/>

結果状態「失明状態」といったように異常状態から起こりえる「結果」属性を得ることができる。原因系の状態変化に関しても同様に「絶対的血糖低下」といった異常状態の「原因」属性を辿ることによって得られる。

なお、現状の臨床医学オントロジーでは、疾患概念について12診療科の疾患が定義されており、このようにして集められた疾患概念の数は合計約6,000となっている。また、その他の概念も含め、疾患オントロジーで定義された概念の総数は約15,000となっている。

## (2) 疾患における異常状態の連鎖の可視化ツールの試作

前項の述べたように、ある疾患の意味を正しく捉えるには、疾患定義で注目した異常状態の変化を含む、連鎖的に引き起こされる一連の状態変化の総体を得る必要があった。そこで4.2.2項と同様の仕組みで、疾患概念のオントロジーから全ての起こりうる異常状態の状態変化を探索し可視化するツールを試作し、疾患概念の定義内容に加えて起こりうる異常状態の状態変化も確認出来るようにした(図8)。図8の画面例では、疾患「大動脈弁狭窄症」の原因である異常状態「大動脈弁口の狭小化」は、その結果状態として「心拍出量の低下」がみられ、「失神・めまい」、「倦怠感」といった状態変化が起こりうることを示している。この可視化ツールでは汎用的な異常状態として個別に定義された概念定義の中から「原因」及び「結果」スロットとして定義された関係性を利用して異常状態の状態変化(因果連鎖)を可視化している。また、このマップ生成では「原因」及び「結果」スロットで記述された状態変化の関係(連鎖関係)を、その方向性を考慮して一方向で表示されるように工夫している。

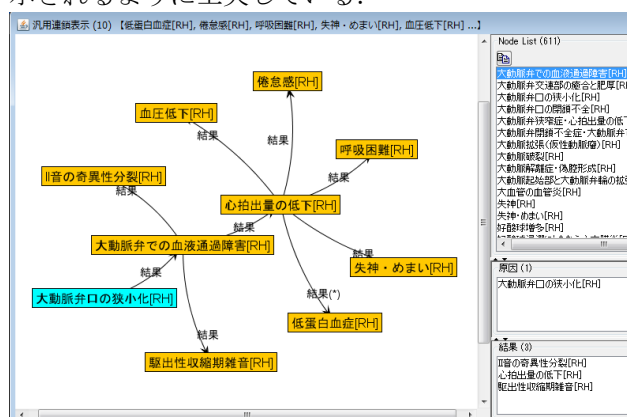


図8 疾患オントロジーの可視化ツールの画面例

## 5. 関連研究

オントロジー構築時における形式的な評価に関し

て、多くのオントロジー構築ツールでは、編集時に間違いを受け付けないことでオントロジーが既読通りに記述されることを保証している。そのため編集時には、利用者間違いのない厳密な記述が要求され、その煩わしさが利用者の発想を阻害してしまう恐れがある。一方法造では、編集時には最小限の整合性検証を行い利用者の発想を阻害しないよう特定のエラーを許容して、オントロジーとしてある程度の品質を保ちつつ、編集時にカバーできない検証項目に関して編集後にオントロジー全体の検証を行うようにしている。このように編集時と編集後に検証を切り分けることで、効果的にオントロジーの形式的な観点での品質保証を実施している。またOWLベースの構築ツールでは整合性検証に記述論理(DL)に基づく推論機構を利用しているが、法造では整合性チェックの推論処理に関しては独自の概念記述を採用している為、そのオントロジーモデルの違いから独自の推論手法を提案している。実際の法造の推論機構ではオントロジーの検証項目に対して個別の推論処理を実現することで高速化を図っている。

オントロジー洗練時における内容的な評価に関して、専門家の知識や経験を通しての判断が必要な為、計算機による洗練支援が難しい問題である。既存研における内容的な評価支援手法としては、専門家がオントロジーの記述内容の理解支援のアプローチとして、オントロジーの可視化によって直感的に理解しやすくする手法の検討[10]や、統制言語を用いて自然言語に近い形で記述させ、それを計算機処理で形式的なオントロジー記述に変換するアプローチがある[11]。またTGVizTab[12]のようにオントロジーの検証や分析の為の拡張として可視化ツールを構築ツールに組み込んだ事例も多くみられる。一方、本提案手法では、洗練時に評価目的に応じてオントロジーから必要な情報を探索し、その結果を概念マップとして提示することで利用者の内容理解を支援することを目指している。その際に可視化が必要となるが、本手法の中心はオントロジーの可視化自体ではなく、利用者の視点に応じた内容を適切に抽出する手法にある。本論文で紹介した適用事例から、何を視点に設定するかは評価目的や対象となるオントロジーに依存するものの、視点に応じたオントロジー探索により概念マップを生成する枠組み自体は汎用的であり、生成された概念マップを用いた内容の評価手法は十分な効果が期待されることが確認できた。

## 6. まとめ

本論文では、オントロジーの構築・洗練過程にお

ける品質保証を支える枠組みとして、オントロジー構築・利用環境「法造」で取り組んでいる内容について述べ、実際のオントロジー構築・洗練過程に適用し、その品質保証に貢献していることを示した。

法造における形式的な評価に関して、編集時に利用者の発想を阻害しないよう最小限の整合性検証を行い、編集後に整合性検証の処理を独自の推論機構を用いて対応することで、効果的にオントロジーを評価し品質保証に貢献していることが分かった。また、法造における内容的な評価に関して、ドメインの専門家による評価を対象として、その目的に応じてオントロジーを探索・可視化した概念マップを生成し、オントロジー構築・洗練を支援する手法について提案し、実際のオントロジー構築での適用事例を用いて考察した。サステナビリティ・オントロジーでの構築・洗練では、ユーザが自由に視点を選択し生成した概念マップを利用することで、既存のオントロジーとの意味的な整合も取りながら確認することができた。臨床医学オントロジーでの構築・洗練では、視点に応じた探索に加え、確認したい内容に合わせて可視化する内容を限定することで、この内容評価手法の有効性を示すことができた。

今後の検討課題として、オントロジー構築プロセスにおける品質保証の為の枠組みとして更なる改良が必要である。具体的には、構築時における形式的な評価に関しては、オントロジーに精通していないドメインの専門家でもオントロジー的に正しく記述できるように適切にガイドする必要がある、ガイドラインの提供やツールでの支援機能の実現が望まれる。洗練時における内容的な評価に関しては、本提案手法のドメインの専門家による評価実験に加え、生成した概念マップから得られた矛盾を修正し、オントロジーに反映するような修正支援の枠組みの検討が必要となってくる。

## 謝辞

本研究の臨床医学オントロジーに関する部分は、厚生労働省、医療情報システム開発普及等委託研究費、「平成 19 - 21 年度医療情報システムのための医療知識基盤データベース研究開発事業」の一環として、東京大学大学院医学系研究科 大江和彦教授、同 今井健助教、大阪大学産業科学研究所 国府裕子特任助教と共同で実施されたものである。

## 参考文献

- [1] B. Smith. et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology* 25, 1251 – 1255, 2007
- [2] Presutti, V., Gangemi, A., 2008. Content ontology design

patterns as practical building blocks for web ontologies. In ER 2008, Proc. of the 27th International Conference on Conceptual Modeling, pp.128-141.

- [3] Masuya, H., Mizoguchi, R., 2009. Toward fully integration of mouse phenotype information. In The 2nd Interdisciplinary Ontology Conference, pp. 35-44. 2008
- [4] Sure, Y., Angele, J., Staab, S., 2003. OntoEdit: Multifaceted Inferencing for Ontology Engineering. In *Journal on Data Semantics, LNCS(2800)* pp.128–152.
- [5] 廣田健, 古崎晃司, 斉藤修, 溝口理一郎: ドメイン知識俯瞰のためのオントロジー探索ツールの開発, 第 23 回人工知能学会全国大会, 2009
- [6] T. Kumazawa, O. Saito, K. Kozaki, T. Matsui, R. Mizoguchi: Toward Knowledge Structuring of Sustainability Science Based on Ontology Engineering, *Sustainability Science, Vol.4, No.1*, 2009.
- [7] R. Mizoguchi, H. Kou, J. Zhou, K. Kozaki, K. Imai and K. Ohe: An Advanced Clinical Ontology, In Proc. of International Conference on Biomedical Ontology (ICBO), pp.119-122, 2009
- [8] 大江和彦, 今井健: 臨床医学知識処理を目指した医療オントロジー開発, *人工知能学会誌 Vol. 25 No. 4* (2010)
- [9] 国府裕子, 周俊, 古崎晃司, 今井健, 大江和彦, 溝口理一郎: 臨床医療オントロジーの構築に関する基礎的な考察, 第 22 回人工知能学会全国大会, 2008
- [10] A. Katifori, C. Halatsis G. Lepouras, C. Vassilakis and E. Giannopoulou: Ontology Visualization Methods—A Survey, *ACM Computing Surveys (CSUR), Surv. 39, 4, Article 10*, 2007
- [11] V. Dimitrova, R. Denaux, G. Hart, C. Dolbear, I. Holt and A. G. Cohn: Involving Domain Experts in Authoring OWL Ontologies, Proc. of 7th International Semantic Web Conference, ISWC-2008, 2008
- [12] Alani, H.: TGVizTab: An ontology visualization extension for Protégé, In Proc. Of Knowledge Capture (K-Cap'03), Workshop on Visualization Information in Knowledge Engineering, Sanibel Island, Florida (2003)

## 付録

### A. 法造におけるオントロジー整合性チェックの検査項目（全 33 項目）

分類	チェック項目	編集時	編集後
基本概念に関する制約 (5)	概念名を表すラベルは、(オントロジー内で)重複しない	○	◎
	ロールホルダーのラベルは、基本概念のラベルを含むすべての概念ラベルと重複しない	◎	◎
	概念間の一般-特殊(is-a)関係について、1つの概念が2つ以上の上位概念を持たない	○	◎
	下位概念として、自身の上位概念を指定できない	○	◎
	本質的な継承関係(is-a関係)を伴わない、弱い継承関係(IS-A関係)は存在しない	◎	◎
スロット(ロール概念, ロールホルダー)に 関する制約 (13)	同一コンテキスト内(=概念)で、スロットのロール名が重複しないこと	—	◎
	概念間のis-a関係が成立する場合に、その概念間のスロット同士でロール名が重複しない	◎	—
	個数制約(カーディナリティ)の表現のフォーマットエラー	◎	◎
	スロットの特殊化を行っている場合の個数制約(カーディナリティ)の数に矛盾がない	—	◎
	設定されている上位スロットがオントロジー内に存在しない	○	◎
	クラス制約からのスロット継承に対して、クラス制約のスロットでないものが設定されている	◎	◎
	上位概念からのスロット継承に対して、上位概念のスロットではないものが設定されている	◎	◎
	上位スロットからの継承に対して、上位スロットのスロットではないものが設定されている	◎	◎
	上位スロットが、上位概念からでもクラス制約からのものでもない	◎	◎
	スロット種別が正しく設定されていない(フォーマットエラー)	—	◎
	上位スロットのスロット種別が異なる(継承されていない)	—	◎
	クラス制約として指定した概念(ラベル)がオントロジー内に存在しない	—	◎
	スロット継承関係に対して、クラス制約の継承関係が成立しない	◎	◎
関係概念に関する制約 (3)	関係概念で定義された関係リンクの数に関して、関係概念のスロットで指定された個数制約を満たすリンク数である	○	◎
	関係概念で定義された関係リンクのスロットに関して、関係概念のスロットで指定されたクラス制約がリンク先のクラス制約として指定された概念と同じ概念あるいは上位概念である	◎	◎
	関係リンク先のスロットが存在しない	◎	◎
データタイプ、基本関係 概念に関する制約 (8)	クラス制約としてデータタイプを指定した場合、制約値として指定された値がデータタイプとしての値の範囲内にある	○	◎
	クラス制約としてインスタンスを指定した場合、制約値として指定された値がインスタンスとして定義されている	—	◎
	基本関係概念(sameAs)の関係リンクに関して、関係概念のスロットで指定されたクラス制約が、リンク先のクラス制約として指定された概念と同じ概念である	◎	◎
	クラス制約がデータタイプである基本関係概念(equal, not-equal)の関係リンクに関して、制約値として指定された値の関係に矛盾がない	◎	◎
	最上位概念であるAny概念が設定されていない	◎	—
	最上位としてAny概念が設定されているにもかかわらず、Any概念を上位概念としない概念が存在する	◎	◎
	スロットの制約値に概念ラベルが指定されている	—	◎
	スロットの制約値のインスタンスがクラス制約条件に合致しない	—	◎
個体(インスタンス)に 関する制約 (4)	定義元概念のクラスおよびオントロジーが指定されている	○	—
	概念クラスの定義に対して、個数制約の違反無くスロットが正しく実装されている	○	—
	個体内の各スロットに対して、クラス制約となる個体が指定されている	○	—
	インスタンスモデル内で定義された関係リンクが定義元概念の関係概念の制約を満たしている(基本関係概念も含む)	○	—

凡例) ◎:新たに追加実装した検証項目 ○:既に実装済み —:編集時(後)に対応のため実装せず