

Is-a 階層の相似性に基づくオントロジー内容洗練支援システムの試作

A Prototype of Ontology Refinement Support System based on Similarity among Is-a Hierarchies

古崎 晃司*¹
Kouji Kozaki

増田 壮志*²
Takeshi Masuda

溝口 理一郎*¹
Riichiro Mizoguchi

*¹ 大阪大学産業科学研究所

*² 大阪大学工学部

The Institute of Scientific and Industrial Research (ISIR), Osaka University School of Engineering, Osaka University

Quality of an ontology is important because it is connected directly with the performance of an application system using the ontology. However ontology refinement to improve its quality needs knowledge and experiments in ontology development. Therefore, ontology refinement task is too difficult especially for beginners in ontology building. In order to solve this problem this article proposes an ontology refinement support system based on similarity among *is-a* hierarchies and an evaluation of it. The system can support content refinements for ontologies.

1. はじめに

オントロジーは対象世界に現れる概念を体系化し、知識システムの基盤を与えるものとして注目され、様々な領域におけるオントロジー構築が進められている。オントロジーの品質は、それを利用する応用システムの能力に直結する重要な要素であり、より良いオントロジーを構築することは、かねてより重要な課題とされている。しかしながら、オントロジー構築には、オントロジーそのものについての知識や経験と、オントロジーを構築しようとしているドメインの専門知識が必要である。そのため、オントロジー構築の初心者にとって、良質なオントロジーを構築することは容易ではなく、オントロジー構築方法やその支援方法の確立が望まれている。一連のオントロジー構築課程の中でも、本研究では、ある程度構築されたオントロジーの品質を向上させるために行う洗練作業に注目する。

オントロジー洗練手法には、オントロジーの文法規則に関する形式的なエラーの有無を検証・修正する方法と、オントロジーの内容を洗練する方法の 2 種類がある。オントロジーの形式的なエラー修正については、推論機構の整合性検証機能を用いた手法が多く提案されている[太田 11, Ohta 11]。一方、オントロジーの内容洗練を対象とした研究は少なく、Voting (投票) システムによるランキングや、概念間の関係性をドメインの専門家の視点に応じた内容の把握しやすい形式で可視化を通じた内容評価の支援 [廣田 09, Kozaki 11] など、属人的な手法の提案にとどまっている。しかし、これらアプローチでは、内容評価にドメイン知識やオントロジー構築に関する知識が不可欠であり、オントロジー内容洗練を実質的に支援しているとは言い難い。

そこで、本研究では、ドメインやオントロジー構築に関する知識の有無に関係なく利用可能な、汎用的な内容洗練手法を考察する。さらに、その洗練手法の適用が可能と思われる箇所をシステムが推定し、修正方法をユーザに提示することで、オントロジー構築の知識や経験に乏しい初心者であっても、ある程度構築されたオントロジーの内容洗練が行えるオントロジー内容洗練支援システムを提案する。本システムは、これまでの研究では困難とされていた、意味的な観点からのオントロジー洗練支援を実現する、世界に類を見ない試みと言える。

以下、2 章では、本研究が目的とする *is-a* 階層の相似性に基づいたオントロジーの内容洗練手法と、その手法に基づいたオ

ントロジー洗練支援システムについて述べる。続く 3 章では提案手法を既存のオントロジー洗練支援に適用した評価実験について述べ、4 章では本研究の現状の総括と今後の課題について述べる。

2. Is-a 階層の相似性に基づくオントロジーの内容洗練手法

2.1 オントロジーにおける *is-a* 階層の相似性

良質なオントロジーを構築するためには、一貫性を持った概念の体系化が求められる。とりわけ、概念間の *is-a* 関係 (一般-特殊関係) に基づく *is-a* 階層は、オントロジーにおける概念体系の基盤を与える最も重要な要素となる。*Is-a* 階層における下位概念は、上位概念をより詳細な (特殊な) 概念に分類したものと位置づけられる。その際、どのような観点で概念が分類されるかは、上位概念から継承された概念定義の内容が下位概念でどのように特殊化されているかに現れる。

例えば、図 1 のオントロジーでは、「乗り物」が移動空間の違いによって「陸上乗り物」と「航空機」に分類されることや、「自転車」がハンドルの違いによって「シティサイクル」と「マウンテンバイク」に分類されることが表されている。ここで、「乗り物」の定義内容を表す移動空間スロットなどは、オントロジー内の他の概念 (移動空間の例では「自然空間」) を参照して定義される。

参照先の概念もそれぞれ *is-a* 階層を持っており、下位概念の定義内容で参照する概念が特殊化される場合には、上位概念の定義内容は全ての下位概念でも満たされるという *is-a* 関係の性質から、下位概念の定義内容におけるスロットの参照先は上位概念の定義内容における同一のスロットの参照先の下位概念とならねばならない。すなわち、*is-a* 階層に沿った概念の定義内容の特殊化は、定義内容のスロットが参照する概念の *is-a*

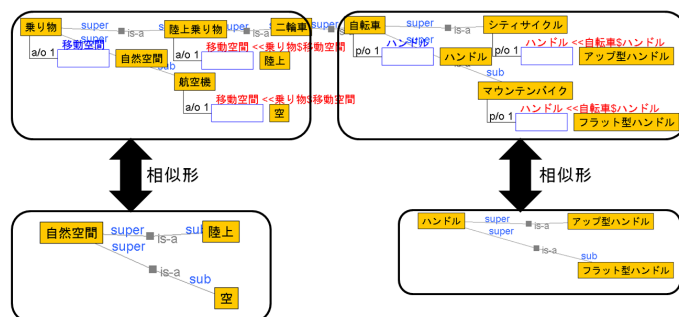


図 1 オントロジーにおける *is-a* 階層の相似性

連絡先: 古崎晃司, 大阪大学産業科学研究所, 〒567-0047
大阪府茨木市美穂ヶ丘 8-1, Tel: 06-6879-8416,
Fax: 06-6879-2120, E-mail: koaki@ei.sanken.osaka-u.ac.jp

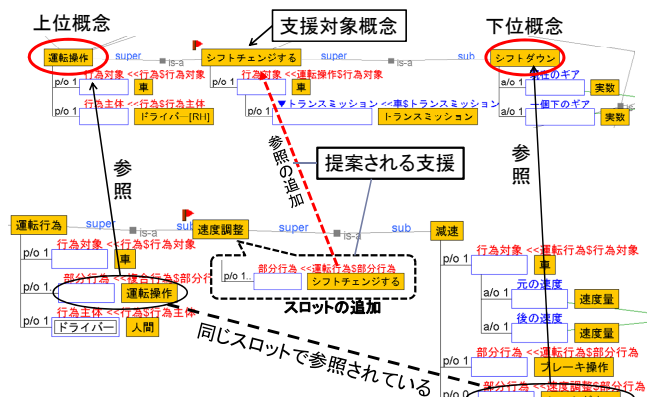


図2 Is-a階層の相似性に基づくオントロジーの内容洗練支援例
階層に沿って行われる。その結果、図1のように、参照元となる概念(の定義内容)の is-a 階層と参照先の is-a 階層には、相似形となる部分が見られる。

このような相似形となる部分はすべての is-a 階層に現れる訳ではない¹。しかし、一貫性を持った is-a 階層を構築するためには is-a 階層に伴う概念分類の観点に一貫性があることが望ましく、参照元と参照先の is-a 階層の分類観点を統一しようとする。上述のような相似形となる部分が多く現れる。本研究では、このような is-a 階層の相似性に基づいて、オントロジーの内容洗練を支援する手法を提案する。

2.2 Is-a 階層の相似性に基づくオントロジーの内容洗練支援

前節で述べたように、良質なオントロジーでは、統一した観点による概念定義が推奨されるため、オントロジー内には、部分的に is-a 階層が相似形となっている箇所が多くみられる。しかし、初心者の構築したオントロジーには、時折参照の抜けがあるため、is-a 階層が相似形になっていない箇所がある。そこで本研究では、このような参照の抜けがあると思われる概念に対して、is-a 階層が相似形になるような修正をオントロジー構築者に提案するといった洗練支援を考える。

Is-a 階層の相似形は、ある概念の定義内容における「参照先の概念の is-a 階層」と、その「参照元の概念の定義内容の is-a 階層²」の間に現れる。よって、オントロジー洗練支援のアプローチとしては、どちらの is-a 階層に合わせて他方の修正を提案するかという方向性の違いから、

- (1) 参照先の概念の is-a 階層と相似形になるように、「参照元の概念の定義内容の修正」を提案する
- (2) 参照元の定義内容の is-a 階層と相似形になるように、「参照先の概念の is-a 階層の修正」を提案する。

という2種類の支援方法が考えられる。本稿では、これらのうち(1)の支援方法について考察する。

例えば、図2のオントロジーでは、「シフトチェンジする」という概念は他の概念から参照されていないが、その上位概念「運転操作」と下位概念「シフトダウン」はそれぞれ「運転行為」および「減速」の定義内容における、同じ「部分行為」スロットで参照されている。このとき、『「運転行為」と「減速」の中間概念にあたる「速度調整」に同じ「部分行為」スロットを追加し、そのスロットで「シフトチェンジする」を参照することで、これらの参照元の定義内容と参照先の is-a 階層を相似形にする』という提案ができる。

¹ 例えば、is-a 階層の途中で異なる分類観点が適用される場合など。

² 厳密には、参照元で定義される「参照先の概念をプレイヤー(クラス制約)とするロール概念(およびロールホルダー)」の is-a 関係を意味する。

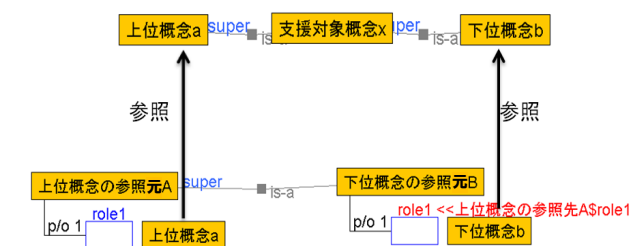


図3 支援適用先候補 (分類1: 上位概念と下位概念が参照されている場合)

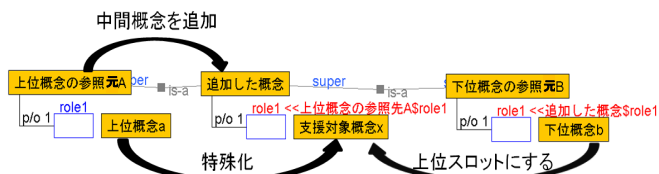


図4 提案される洗練支援(分類1)

このような is-a 階層を相似形にするようなオントロジーの修正支援が提案できる箇所の候補について考察すると、いくつかのパターンに分類することができる。そのようなパターンの計算機で判定可能な分類条件を明示化することにより、is-a 階層の相似性に基づいたオントロジーの内容洗練支援を適用できる箇所の候補を、機械的に推定することが可能となる。

次節では、その支援適用先候補の分類について考察する。

2.3 支援適用先候補の分類

2.2 節で述べた(1)の支援方法では、「参照先の概念の is-a 階層」と「参照元の定義内容の is-a 階層」が相似形となっていない箇所のうち、前者の is-a 階層は変更せずに、後者を修正することで両者が相似形になるような提案ができる箇所が支援対象となる。形式的な記述エラーを含まないオントロジー³において、「他のいずれの概念の定義内容からも参照されていない概念」は、常にこのような支援対象の候補となる条件を満たす⁴。よって本研究では、他の概念から参照されていない概念を「支援対象概念」とする。

このような支援対象に対して、is-a 階層を相似形とするために、どのような修正を提案できるかは、支援対象概念と is-a 関係にある概念の存在の有無や、それらの概念を参照している概念の定義内容の有無、それらの関係性などの組み合わせによって決まる。ここでは、そのような組み合わせの一般形について考察し、支援適用先候補を分類する。

例えば図3のように、「支援対象概念の上位概念および下位概念が共に、他の概念の定義内容から同じスロットで参照されており、それらの参照元となる概念およびスロット間に is-a 関係が成立している場合」を考える。このとき、is-a 階層を相似形にするためには、『参照元となる概念間に中間概念を追加⁵し、その中間概念の定義内容の同じスロット⁶で支援対象概念を参照する』という修正支援(図4、表1の「支援方法1」に相当)を提案できる。一方、図3と同様に支援対象概念の上位概念と下位概念が共に参照されている場合、図2のように参照元の概念間に

³ 形式的な記述エラーは推論機構による整合性検証によって検出可能なため、あらかじめ修正されていることを前提としている。

⁴ 他の概念から参照されている概念であっても、参照のされ方によっては、支援対象の候補となり得るが、その詳細な考察は今後課題とする。

⁵ 追加する中間概念のラベル(概念名)はユーザが決定する必要がある。

⁶ 中間概念の定義内容に「同じスロット」を追加するには、「中間概念の定義内容において、上位概念から継承されたスロットを特殊化した上で、下位概念の同名スロットの上位スロットとする」という操作が必要となる。

中間概念存在する場合は、先の例で述べた修正支援に加え、『その元から存在している中間概念の定義内容の同じスロットで、支援対象概念を参照する』という修正支援(表 1 の「支援方法 2」に相当)も提案することができる。

さらに、支援対象概念の上位概念のみ/下位概念のみが他の概念から参照されている場合を考慮すると、支援対象概念に対して is-a 階層を相似形にするための修正支援を提案できるパターンは、表 1 に示した 6 種類に分類できる。紙面の都合で詳細は割愛するが、これら 6 種類のパターンへ重複無く分類するための適切な分類条件は既に明らかになっている。

なお、これらのパターンでは、支援対象概念の直上の上位概念/直下の下位概念しか考慮しておらず、2 段以上の上位/下位概念や、兄弟概念が考慮されていない。しかし、これらの場合についても、6 種類のパターンを繰り返し適用することで、is-a 階層を相似形にするような修正案を推定できる。このような複数回適用も含めると、他の概念から参照されていない概念を「支援対象概念」とした際に、支援が提案できるパターンはこれら 6 種類で網羅されている。

ただし、提案された修正を適用することが、オントロジーを洗練する上で、必ずしも適切であるとは限らない。そのため、提案された修正を適用する順番によっては、すべての支援対象概念に対して洗練支援適用先候補が推定されるとは限らない。

2.4 オントロジー内容洗練支援システムの試作

提案した洗練手法に基づきオントロジーの内容洗練支援システムを設計・試作した。本システムは、2.3 節で述べたパターンに基づき修正支援を適用する箇所の候補と提案する修正内容を推定する「支援適用先候補推定モジュール」、推定された内容に基づき洗練支援案をユーザに提示し、適用する修正案を選択するための「支援適用先候補表示・選択モジュール」、ユーザが選択した修正案を適用し、オントロジーの定義内容を修正する「支援適用モジュール」から構成される。なお、本システムの実装には Java を用いており、オントロジーを処理するためのライブラリとして HozoCore および法造 OAT (Ontology Application Toolkit)を利用している¹⁾。

3. 洗練支援手法の評価

3.1 評価方法

3 章で提案したオントロジーの内容洗練支援手法の有用性を評価するために、試作した内容洗練支援システムを実際に構築されたオントロジーに適用する評価実験を行った。

評価実験の対象としたオントロジーは、初心者が構築したオントロジー-9 つ²⁾と、オントロジー構築の熟練者が構築し一般に公開しているオントロジー-3 つ³⁾とし、以下の手順で行った。

(1) 提案手法の適用範囲の評価

各オントロジーについて、

- a) オントロジー中の全概念数に対する、提案手法の支援対象概念となる「他の概念から参照されていない概念」の割合(支援対象概念の割合)

1 HozoCore および法造 OAT は、<http://ontsupport.enegate.jp/ontology/>にて公開されている。

2 初心者が初めて構築したオントロジー5 つと、オントロジー構築の熟練者のアドバイスを受けて、それらを修正したもの4 つ。

3 http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm YAMATO および、法造 (<http://www.hozo.jp>) のサンプルファイル vehicle.xml と soccer_ver.12.xml

表 1 修正支援適用先候補の提案パターンの分類

	分類 1	分類 2	分類 3	分類 3	分類 4	分類 4	分類 5	分類 5	分類 6	分類 6
上位概念が参照されている	○	○	○	○	○	○	○	×	○	×
下位概念が参照されている	○	○	○	×	○	×	○	○	○	○
下位概念を参照しているスロットが上位概念を参照しているスロットの継承スロットである	○	○	×	/	×	/	×	/	×	/
上位概念と下位概念を参照しているスロットが直接 is-a 関係である	○	×	/	/	/	/	/	/	/	/
上位概念を参照している概念の下位概念の有無	○	○	○	○	×	×	-	-	-	-
下位概念を参照している概念の上位概念が any では無い	○	○	-	/	-	/	×	×	○	○
支援方法 1	○	○	○	○	○	○	○	○	○	○
支援方法 2	×	○	○	○	×	×	×	×	○	○

凡例： ○…条件を満たす、×…条件を満たさない、----Don't care、/…考慮できない

- b) 全支援対象概念の数に対する、支援適用先候補を推定することができた概念の数の割合(支援適用候補先提示率)
- c) 推定された支援適用先候補の分類ごとの割合を調べる。

(2) 推定された支援適用先候補の妥当性の評価

初心者が構築したオントロジーのうちの 1 つに対しシステムが推定した支援適用先候補について、「提案された修正方法を適用することが、オントロジーの内容洗練として適切かどうか」を実験者が判断し、適切と思われる提案の割合を調べる。

3.2 実験結果と考察

(1) 提案手法の適用範囲

表 2 に、実験結果を示す。支援適用候補先提示率を見ると、提案手法の適用範囲としては、初心者が構築したオントロジーについては 49%、熟練者が構築したオントロジーについては 34%の支援対象概念に対して洗練方法を提案することができたことが分かる。なお、洗練方法を提案することができなかった支援対象概念は、2 段以上の上位/下位概念や兄弟概念のみが他の概念に参照されているものであるが、これらについても提案された洗練方法を繰り返し適用することで、洗練支援方法を提案することができる。

次に推定された支援適用先候補の分類ごとの傾向を見ると、支援対象概念の上位概念と下位概念の両方が他の概念から参照されているというより厳しい条件を満たしている分類 1,2 の割合が低いことが分かる。これらは分類の条件が厳しい分、is-a 階層を相似形にする妥当性が高いと考えられる。また分類 5 の割合も低い、その理由は、分類 5 では支援対象概念の下位概念がそのオントロジー内で最上位の概念から参照されていることが分類条件となっているが、最上位の概念は通常はごく少数であるためと考えられる。

(2) 推定された支援適用先候補の妥当性の評価

初心者が構築したオントロジーのうち race2 に関して支援適用先候補の妥当性を確認した結果、20%の候補については提案された修正を適用することが妥当であると判断された(表 3)。

表 2 提案手法の適用範囲の評価結果

	初心者が構築したオントロジー										熟練者が構築したオントロジー			
	race1	race2	drums	drums2	traffic	traffic2	rail	rails2	hero_1	平均	soccer	vehicle	YAMATO	平均
分類1	1	2	1	1	6	7	2	5	0		6	1	10	
分類1/全支援候補	1%	1%	1%	1%	4%	4%	3%	4%	0%	2%	3%	2%	2%	2%
分類2	2	9	2	2	0	10	3	5	0		0	3	24	
分類2/全支援候補	3%	6%	2%	2%	0%	6%	4%	4%	0%	3%	0%	5%	6%	4%
分類3	32	44	23	23	81	80	27	62	67		68	32	148	
分類3/全支援候補	42%	32%	24%	27%	55%	46%	38%	47%	48%	40%	39%	50%	36%	42%
分類4	22	34	40	30	20	23	14	19	68		39	6	120	
分類4/全支援候補	29%	24%	43%	36%	14%	13%	20%	14%	49%	27%	23%	9%	29%	20%
分類5	3	1	0	0	1	1	0	0	1		0	0	4	
分類5/全支援候補	4%	1%	0%	0%	1%	1%	0%	0%	1%	1%	0%	0%	1%	0%
分類6	17	49	28	28	39	53	25	42	4		60	22	104	
分類6/全支援候補	22%	35%	30%	33%	27%	30%	35%	32%	3%	27%	35%	34%	25%	31%
集計														
概念の全数	133	212	135	135	193	251	112	175	50		261	122	561	
全参照無し概念数	79	88	58	58	87	69	77	97	29		148	75	363	
支援適用先推定可能概念数	34	52	34	34	40	41	29	36	11		55	28	101	
支援適用先候補の総数	77	139	94	84	147	174	71	133	140		173	64	410	
支援対象概念の割合	59%	42%	43%	43%	45%	27%	69%	55%	58%	49%	57%	61%	65%	61%
支援適用先候補提示率	43%	59%	59%	59%	46%	59%	38%	37%	38%	49%	37%	37%	28%	34%

※同一の支援対象概念に対して、複数の支援適用先候補が提案されることがあるため、「支援適用先推定可能概念数(少なくとも1つの支援適用先候補が推定された支援対象概念の数)」と「支援適用先候補の総数」は一致しない。

表 3 提案された洗練支援の適用妥当性の評価結果

	上下		上位のみ		下位のみ		合計
	分類1	分類2	分類3	分類4	分類5	分類6	
提案された支援適用先の数	2	9	44	34	1	49	139
適用が妥当と思われる支援の数	2	5	2	8	0	11	28
適用が妥当な数の割合	100%	56%	5%	24%	0%	22%	20%

提案された修正の適用が不適切であると判断された例としては、支援対象概念が「人工物」であるとき、その上位概念「実在物」が「行為」の対象物スロットで参照されていることを考慮し、『「行為」の下位概念として対象物が「人工物」である「人工物対象行為」といった概念の追加を提案する』といったもので、この提案の適用によって追加される「人工物対象行為」¹は不自然であるため不適切であると判断された。このように、トップレベルに近い上位概念を参照した概念定義は、オントロジーの定義内容としては誤りではないが、「人工物対象行為」のように日常的には概念化されていないため不自然に思われる場合が多い。本評価実験で提案された修正の適用が不適切と判断されたものには、これと同様な例が多く含まれていた。

そこで、支援対象概念からトップレベルに近い概念を除いたところ、表 4 に示すように、提案された修正を適用することが妥当であると判断される割合が 39%に向上した。

3.3 提案手法の有用性の考察

本手法では、提案された修正を実際の洗練作業に適用するか否かの判断はオントロジー構築者の判断に委ねられるので、提案の妥当性が 39%であれば、十分な有用性があると思われる。しかも、また、適用が不適切と判断された修正の提案についても、前述のようにオントロジーの定義内容としては誤りで無く、日常的な概念化とのギャップによる不自然さが主な問題となるだけであり、オントロジー洗練作業を大きく妨げるものではない。

しかしながら、推定された支援適用先候補の妥当性の評価は、初心者が構築したオントロジーの 1 つに対してのみしか実施しておらず、対象とするオントロジーの数を増やし、より詳細な評価と考察を重ねる必要がある。特に、本システムの有用性を高めるには、提案された洗練支援を適用する優先順位の決定方法や、適用順番を変えたときの繰り返し適用後に推定される支援適用先候補への影響に関する考察は、特に重要と思われる。

¹ 追加される概念のラベル(概念名)はユーザが決定すべきであるが、対象物が「人工物」である行為であるので、ここでは仮に「人工物対象行為」と呼ぶ。

表 4 提案された洗練支援の適用妥当性の評価結果 (トップレベルの概念を支援対象概念から除いた場合)

	上下		上位のみ		下位のみ		合計
	分類1	分類2	分類3	分類4	分類5	分類6	
提案された支援適用先の数	2	5	25	18	0	21	71
適用が妥当と思われる支援の数	2	5	2	8	0	11	28
適用が妥当な数の割合	100%	100%	8%	44%	0%	52%	39%

る。また、is-a 階層の相似性に基づく洗練支援の適用範囲の限界についての考察も重要である。表 2 で熟練者が構築したオントロジーにおいても、支援対象概念が多くあることから、is-a 階層の相似性が成立しないのはどのような場合かを、より深く検討する必要が示唆されている。

4. まとめ

本研究では、オントロジーの中には部分的に is-a 階層が相似形となる箇所が現れることに着目し、is-a 階層の相似性に基づくオントロジー内容洗練支援システムを提案し、その有用性の評価を行った。本手法の特徴は、機械的な支援が困難であると思われていたオントロジーの内容洗練を限定的な範囲ではあるが実現している点にある。今後、本手法を発展させることで、より広範囲で精度の高い、洗練支援の実現が期待される。

謝辞

本研究の一部は科学研究費補助金若手研究(A) 20680009 の助成による。

参考文献

[Kozaki 11] Kouji Kozaki, et al.; Understanding an Ontology through Divergent Exploration In Proc. of 8th Extended Semantic Web Conference (ESWC2011), pp.305-320, Heraklion, Greece, May 29 - June 2, 2011.

[廣田 10] 廣田 健, 古崎 晃司, 齊藤 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, 人工知能学会第 23 回全国大会(JSAI2009), 213-2, 2009

[Ohta 11] Mamoru Ohtai, et al.: A Quality Assurance Framework for Ontology Construction and Refinement, In Proc. of 7th Atlantic Web Intelligence Conference (AWIC2011), pp.207-216, Fribourg, Switzerland, January 26-28, 2011.

[太田 11] 太田 衛, 古崎 晃司, 溝口 理一郎: 実践的なオントロジー開発に向けたオントロジー構築・利用環境「法造」の拡張 — 実践編 —, 人工知能学会論文誌, Vol.26 No.2, pp.403-418, 2011.