

兄弟概念の分類観点に着目したオントロジー一貫性向上手法の開発

Development of System to Improve Consistency of Classification among Brother Concepts

増田 壮志*¹
Takeshi Masuda

古崎 晃司*¹
Kouji Kozaki

駒谷 和範*¹
Kazunori Komatani

*¹ 大阪大学産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University. *¹

The consistency of classification is one of characteristics that indicate the quality of ontologies. We focus on the consistency among brother concepts that have a same parent concept. In this paper, we develop a refinement system that finds inconsistent parts from a target ontology and propose methods to make such parts consistent. To find inconsistent parts and make proposals, the system compares a slot hierarchy and other two hierarchies that have reference relationship to the slot hierarchy.

1. はじめに

オントロジーは、医療情報、環境問題や生物規範工学等の様々な分野で構築され、知識システム開発の基盤として用いられている。しかしながら、品質の高いオントロジーの構築は容易ではないため、構築方法論に関する研究や、構築支援のためのシステムの開発が進められてきた。本研究では、既存のオントロジーの品質を高めるような構築支援システムの開発を行う。

オントロジーの洗練段階における構築支援は、文法規則のエラー修正と定義内容の洗練の2種類に大別できる。前者は、多くのオントロジー構築ツールで実装されている入力支援や文法チェック機能、オントロジーの推論機構を用いた整合性検証機能[Sirin 07, Kopena 03, Decker 98 太田 10]等がある。一方、後者の定義内容の洗練を対象としたアプローチとしては、オントロジーの分類上の関係の妥当性を強化する OntoClean[04] やドメイン知識俯瞰のためのオントロジー探索ツールの利用[M. Ohta 11]などがある。

しかしながら、既存の構築支援システムでは、文法エラーの検出、修正は可能であるが、定義内容の洗練・評価に関しては、対象ドメインの専門家への提示方法の工夫にとどまっており、品質向上のための具体的な洗練箇所及び方法を提示することはできない。そこで本研究では、オントロジーの構造情報を元にした概念階層の比較によって、洗練箇所と洗練方法を検出するシステムを開発する。概念階層比較方法には、上位下位の関係に着目した概念階層比較方法と、兄弟概念の関係に着目した概念階層比較方法の2種類ある。筆者らはこれまでの研究で、上位下位の関係に着目した概念階層比較方法の開発を行い、その有効性を確認した[増田 15]。しかし、上位下位の概念階層同士を比較する場合、検出される洗練箇所の兄弟概念の有無は考慮されないため、システムユーザは、基本概念階層や被参照概念階層の兄弟概念がどのような分類観点で特殊化されているかわからない。そのため、提案された概念及びスロットの追加が、兄弟概念と一貫性があるかの判断ができないため、提案された洗練方法の正否の判断が困難であるという問題があった。

本論文では、そうした問題の解決のために開発した兄弟概念の関係に着目した概念階層比較方法について述べると同時に、兄弟概念に着目した手法で新たに発生した洗練方法数の膨張への対応についても述べる。

連絡先: 増田 壮志, 大阪大学産業科学研究所 知識科学研究分野, 〒567-0047 大阪府茨木市美穂ヶ丘 8-1, Tel:06-6879-8416, e-mail:masuda@ei.sanken.osaka-u.ac.jp

2. 概念階層の比較によるオントロジー一貫性向上

本研究では、構築されたオントロジーに対して、概念分類の一貫性が崩れている箇所を洗練の必要な箇所として検出し、洗練提案を行う。そこで一貫性が崩れているどうかの指標として「ある概念の下位概念において、その分類は同じ性質を参照して行われることが望ましい」[溝口 06]というオントロジーの構築指針に着目した。この指針に沿って構築されたオントロジーには、参照関係にある3つの is-a 階層が互いに相似形となるという性質が現れる(図 1)。参照関係にある3つの is-a 階層とは、(1)概念定義の際に各概念の定義内容を表し概念の特殊化に伴って特殊化されるスロットの階層、(2)それらのスロットを持つ概念の階層(基本概念階層)、および、(3)スロットから参照されている概念階層(被参照概念階層)の3つの階層である。本研究では、これら3つの概念階層をスロット階層を中心に比較し、非相似となっている部分を洗練箇所として検出し、相似となるような洗練提案を行う。

3. 兄弟概念階層比較による洗練支援

3.1 兄弟概念階層比較による洗練箇所の検出

本論文で述べる兄弟概念を含む概念階層の比較手法においても、2章で述べた通り、基本概念階層、スロット階層、被参照概念階層の3つの is-a 階層の比較によって洗練箇所の検出を行う。以前の手法では、上下の概念階層同士を比較していたのに対し、本論文で述べる手法では、兄弟概念も含めた概念階層同士を比較するため、兄弟概念の分類観点を考慮しながら洗練提案の適用の可否を判断することができる。

兄弟概念も含めた概念階層の比較手法は次のような流れで行われる。あるスロット S とその一段下のスロット全て SL_m ($m=1\sim M$, $M:S$ の一段下のスロットの総数) を比較元となるスロット階層として、それらのスロットを持つ概念とその兄弟概念からなる基本概念階層と、スロット階層から参照されている概念とその兄弟概念からなる被参照概念階層を比較する。この時基

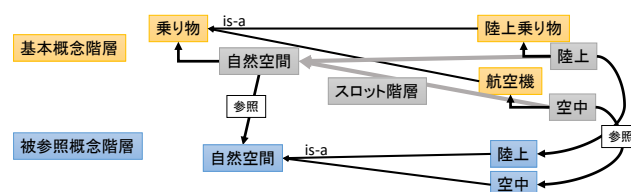


図 1 相似形な is-a 階層

表 1 兄弟概念の有無による洗練箇所分類

		基本概念階層の兄弟概念数 B_BR _x	
		X>0	X=0
被参照概念階層中の兄弟概念数 R_BR _y	Y>0	(i) 非相似	(ii) 非相似
	Y=0	(iii) 非相似	(iv) 相似

本概念階層中のスロットを持つ概念は、スロット階層のスロットと対応するように B, BL_mと表現し、被参照概念階層中の参照されている概念も同様に、R, RL_mと表現する。また、BとBL_mの中間概念を BM_a(a = 1~A, A: 中間概念の総数), RとRL_mの中間概念を RM_b(b = 1~B, B: 中間概念の総数)とする。基本概念階層及び被参照概念階層は、着目スロット群と参照関係にある概念の他に兄弟概念にあたる部分も含まれるが、これらの兄弟概念は、基本概念階層では、B_BR_x(x=1~X, X: 兄弟概念の総数)、被参照概念階層では、R_BR_y(y=1~Y, Y: 兄弟概念の総数)と表現する(図 2)。今回の手法では、上位下位の概念階層比較手法[増田 15]に相当するため BM_a及び RM_bに概念及びスロットの追加を行わない。

このように兄弟概念を含む概念階層を比較することによって、着目しているスロットを持つ概念の兄弟、参照されている概念の兄弟を合わせて洗練箇所として検出、提示することが可能となったので、以前の手法よりも多くの情報から提案の可否を判断できるようになった。例えば、「B_BR₁でRL₂を参照したスロットを追加する」といった提案がなされた場合、破線部だけの提示であれば、B_BR₁とRL₂の2つの概念の情報のみでスロットの追加を判断しなければならない。一方、兄弟概念を合わせて提示した場合、図2のすべての概念が提示されるので、B_BR₁には、BL₂という兄弟概念が存在し、BL₂は、追加提案されたスロットの兄弟にあたるスロットを持つことも確認できる。このように兄弟概念階層比較による洗練箇所の検出を行うことで、提示できる概念階層が広がり、洗練提案の適用時にユーザが兄弟概念の分類観点の一貫性を保ちながら適用の判断ができるようになった。

3.2 検出される洗練箇所の分類

このように3階層を比較した場合、基本概念階層及び被参照概念階層に着目スロット群(S_m)と参照関係に無い兄弟概念(B_BR_xまたはR_BR_y)が存在した場合、それぞれのis-a階層は相似形ではない。よってそのような個所を、概念分類の一貫性が崩れていると判断することができる。すなわちB_BR_xとR_BR_yの有無(X, Yの数の組み合わせにより4種類に分類できる(表1)。(i)~(iii)については、概念およびスロットの追加により、参照関係にある概念階層が相似形となるような洗練提案を

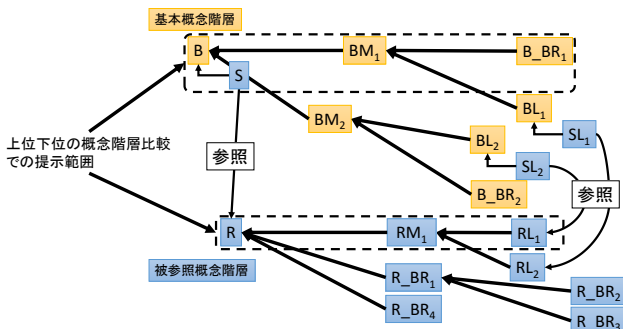


図 2 兄弟概念を含めた概念階層の比較

行う。また、(iv)は、X = Y = 0であり、3つのis-a層が相似形であるため、洗練箇所として検出しない。

3.3 提案される洗練方法

提案される洗練方法は、大まかには次の3種類である。

- スロット追加
- 基本概念階層への概念追加+スロット追加
- 被参照概念階層への概念追加+スロット追加

本研究では、基本概念階層と被参照概念階層の両方への概念追加は提案しない。なぜなら、両方の概念階層への概念追加は、既存の概念階層とは関係なく行うことができるため、際限なく概念追加が提案可能となってしまうからである。

さらに、このうち(a), (b)は、以下の(1)スロットの追加先及び(2)スロットの参照先の選択方法によってさらに細分化される。

(1) スロットの追加先の選択方法

スロットは、基本概念階層中のいずれかの概念に追加することになるため、次の2パターンが考えられる。

- スロットを持たない兄弟概念(B_BR_x)
- 新しく基本概念階層に追加された概念

このとき、新規に追加される概念は、BL_mの兄弟概念とし、BL_mの上位または下位には概念を追加しない。BL_mの上位下位への概念追加は、以前手法[増田 15]に相当する。

(2) スロットの参照先の選択方法

また、スロットは被参照概念中の概念を参照して定義される。この場合の選択肢は、次の3パターンである。

- スロットから参照されていない兄弟概念(R_BR_y)
- 既に兄弟スロットで参照されている概念(RL_m)
- 新しく被参照概念階層に追加された概念

(1)と異なり、既存の兄弟スロットで参照されている概念(RL_m)が含まれているのは、スロットから参照される概念は、重複して選択された場合でも、基本概念階層の分類の観点の混在することがないからである。また3は、(1)と同様に、RL_mの兄弟概念に限定される。

以上、(1), (2)より、概念の追加の有無、スロットの追加先および参照先から洗練方法は表2のように分類でき、それぞれの洗練方法を適用可能な洗練箇所分類は、表2のような対応関係となっている。

3.4 兄弟概念階層同士の比較の問題点とその解決案

(1) 兄弟概念階層を比較した場合の問題点

兄弟概念を含む概念階層同士の比較の際には、上位下位概念階層の比較手法よりも比較対象となる概念数が多いため提案される洗練方法の数が膨大になるという問題点がある。例として、スロットSの下位スロット数M = 2、基本概念階層中のスロットを持たない兄弟概念B_BR_xの数X = 40(この内末端の概念が18個)、被参照概念階層中の参照されていない兄弟概念R_BR_yの数Y = 3(この内末端の概念が2個)のときを考える。この場合X>0, Y>0なので洗練箇所としては分類(i)である。表2より分類(i)の場合、表2中の全ての洗練方法が提案される。よって、提案される洗練方法の総数は、各数値を代入すると計

表 2 洗練方法と洗練箇所の関係

洗練方法	3.3.(1) スロット追加先 (基本概念階層)		3.3.(2) スロット参照先 (被参照概念階層)		3.2. 適用可能な洗練箇所の分類		提案される洗練方法数
	1. スロットを持たない兄弟概念 Bn_BRx	2. 新規追加概念	1. Rn_Bry	2. Rn_Lm	分類(i), 分類(ii), 分類(iii)	分類(i), 分類(ii), 分類(iii)	
a 1	1. スロットを持たない兄弟概念 Bn_BRx	2. 新規追加概念	1. Rn_Bry	1. Rn_Lm	分類(i)	分類(i), 分類(ii), 分類(iii)	X*Y
a 2			2. Rn_Lm	2. Rn_Lm	分類(ii), 分類(iii)	分類(i), 分類(ii), 分類(iii)	X*M
b 1	2. 新規追加概念	2. 新規追加概念	1. Rn_Bry	1. Rn_Lm	分類(i), 分類(ii)	分類(i), 分類(ii), 分類(iii)	(Xのうち末端でないもの数)*Y
b 2			2. Rn_Lm	2. Rn_Lm	分類(i), 分類(ii), 分類(iii)	分類(i), 分類(ii), 分類(iii)	(Yのうち末端でないもの数)*M
c 1	1. スロットを持たない兄弟概念 Bn_BRx	3. 新規追加概念	3. 新規追加概念		分類(i), 分類(ii)	分類(i), 分類(ii)	(Yのうち末端でないもの数)*X

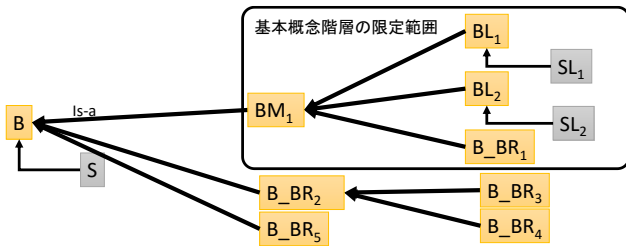


図3 基本概念階層の限定範囲

328 個の洗練方法が提案される。この例では、被参照概念階層の兄弟概念数 $Y=3$ と少なかったが、これが X と同様の数の場合もあり、そうした場合であれば、1 か所あたり数千個の洗練方法が提案されることになる。

(2) 兄弟概念階層の比較対象となる概念階層の限定方法

このような膨大な洗練方法の提案を一度に行うことは、非現実的であるため、これらの洗練方法のうち、どの方法を優先的に表示するかを考える必要がある。提案される洗練方法を限定するには、比較対象となる兄弟概念数(X, Y)が減少すればよい。兄弟概念数が減少すれば、洗練提案の組み合わせが減少するからである。本論文では、次の 2 つの方法で比較対象となる兄弟概念を限定することで提案される洗練方法を減らすことを試みた。

方法 1: 同じ親を持つ兄弟概念に限定

一般的に基本概念及び被参照概念に多くの兄弟概念が存在する場合、複数段にわたる特殊化のために各段での概念の分岐が増え、兄弟概念が増える原因となっていることが多い。そこで、概念階層を同じ親概念を持つ兄弟概念に限定するという方法を考えた。

図 3 に、基本概念階層での例を示す。BL₁ の同じ親を持つ兄弟概念に限定した場合、図 3 中の実線で囲まれた部分を比較することになり、比較対象になる兄弟概念の数を減らすことが可能である。被参照概念階層においても同様に RL₁ と同じ親を持つ兄弟概念に限定する。

方法 2: 親概念から等距離にある兄弟概念に限定

2 つ目は、比較対象となる is-a 階層を親概念から等距離にある、つまり、特殊化の段数が等しい兄弟概念に限定する方法である。

図 4 に例を示す。この時 BL₁ が、B から 1 段特殊化されているので、 $B_{1_length} = 1$ (B_{m_length} : B と BL_m の特殊化の段数) である。比較対象となる兄弟概念は、特殊化の段数が 1 である BL₂、B_BR₁、B_BR₃ の 3 つの概念である。B_BR₂、B_BR₄、B_BR₅ は、 $B_{BR2_length} = B_{BR4_length} = B_{BR5_length} = 2 \neq 1$ であるので比較対象から除外される(B_{BRx_length} : B と B_BR_x の段数)。被参照概念階層も同様に兄弟概念階層を限定する。

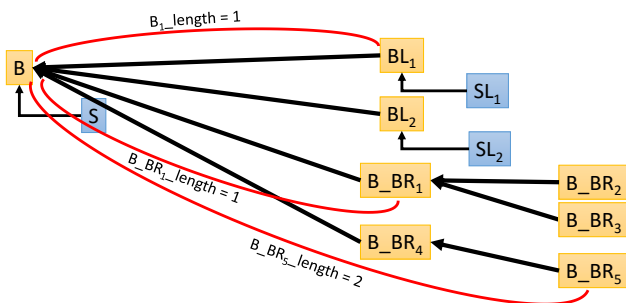


図4 親概念からの特殊化の段数

4. 評価実験

4.1 実験目的・概要

本論分で述べてきた兄弟概念階層比較による洗練支援手法の評価を行う。この評価実験は、本格的な評価実験の前段階として筆者が行った予備実験である。実験には筆者が本研究を始める前にオントロジー構築技術取得のために構築したレースオントロジーを用いた。このオントロジーの総概念数は 213 個である。評価観点は、(1)兄弟概念階層比較によって新規に検出できた洗練箇所数及び適切な洗練提案ができた箇所割合、(2)比較対象とする兄弟概念を限定した場合に(1)で適切な洗練提案ができた箇所をどの程度カバーできるのか、の 2 点である。

このうち兄弟概念を全て比較する方法によって、11 箇所が洗練箇所として検出され、そのうち 7 箇所に対して適用すべき洗練方法が提案された。この 7 箇所に兄弟概念を限定した洗練提案を行う。このときに何ヵ所に対して、限定しない場合に適用すべきとされた洗練方法を提案できたかを評価した。

4.2 実験結果及び考察

(1) 兄弟概念階層比較と以前手法の比較結果

表 3 に以前手法との比較結果を示す。以前手法では、149 箇所の洗練箇所が検出され、そのうちの 18% に対して、適切な洗練提案が可能であった。一方、今回の手法では、11 の洗練箇所を新規に検出することが可能であり、そのうち 63% に対して、適切な洗練提案が可能であった。しかし、以前の手法での 1 箇所あたりの平均提案数が 7 個に対し、1154 個と非常に多い。

(2) 兄弟概念を限定した場合の洗練箇所のカバー率の変化

(1)では、新手法の方が正解率は高いが、3.4.(1)で述べたように 1 箇所あたりの洗練提案が非常に多いという結果となった。次に、(1)で適切な提案ができた 7 箇所について兄弟概念を限定し、提案数を絞った場合にどの程度正解とされた正解箇所をカバーできるのかについて調べる。表 4 に実験結果を示す。

同じ親概念を持つ兄弟概念に限定した場合、平均提案数が 1 箇所あたり 13 個と 100 分の 1 程度にまで減少した。一方、7 箇所のうち 4 箇所に対しては、兄弟概念を限定しても正しい洗練提案を行うことができた。残りの 3 箇所については、同じ親概念を持つ兄弟概念に限定した際に、比較対象となる兄弟概念がなくなってしまい、洗練提案が不可能となった。

次に親概念から等距離にある兄弟概念に限定した場合の結果について述べる。平均提案数は、377 個と 3 分の 1 程度になったとはいえ多いままであった。この時、正しい洗練提案が可能であったのは 7 箇所中 5 箇所であり、残り 2 箇所については同じ親概念を持つ兄弟概念に限定したときと同じ理由により洗練提案不可能となった。

以上の結果より、兄弟概念を限定した場合でも、兄弟概念をすべて比較した場合の洗練提案のうち半数以上に対して、正しい洗練提案が可能となることが確認できた。つまり、以前手法に比べて正解率の高く適用の際の判断が容易な洗練提案が、同程度の提案数で可能となることが分かった。また、洗練不可になってしまった箇所については、少なくとも一つ兄弟概念が出てくる段数まで比較対象とする概念階層を拡張するなどの工夫が必要である。

表 3 以前手法との比較結果

	検出数	適切な提案ができた洗練箇所数	適切な提案ができた洗練箇所割合	平均提案数
上位下位概念を比較する手法	149	27	18%	7
兄弟概念階層を比較する手法	11	7	63%	1154

表 4 兄弟概念を限定した結果

	洗練箇所数	適用すべき提案 ができた 洗練箇所数	洗練提案 ができなくなった 洗練箇所数	平均提案数
①全ての兄弟概念	7	7	0	1154
②同じ親を持つ 兄弟概念に限定	7	4	3	13
③親概念から等距離の 兄弟概念に限定	7	5	2	377

4.3 洗練提案の実例

(1) 比較対象を制限した場合でも洗練提案可能であった例

図 5 に比較する兄弟概念階層を同じ親概念を持つ兄弟概念に限定した例を示す。ここでは、基本概念階層は、破線部分に限定され、被参照概念階層は、実線部分に限定される。被参照概念階層に限定範囲が 2 か所存在するのは、“減速”において 2 つの異なる親を持つ概念が参照されているためである。被参照概念階層の限定範囲 1 では、“アクセル操作”を“加速”で参照するといった提案がされ、限定範囲 2 では、“シフトアップ”を“加速”で参照するといった提案がされる。この場合どちらの洗練提案も適用が妥当である。さらに限定範囲毎に提案される洗練方法は、先ほど述べたものを含め 2 つのみであり、限定範囲 1, 2 で計 4 つである。この洗練箇所にはもともと 288 個の洗練提案がなされていたので、多数の提案を適切なものに限定することができている。

(2) 比較対象を制限した場合に洗練提案できなくなった例

比較対象を同じ親概念を持つ兄弟概念に限定した場合に、提案不可能となった例を図 6 に示す。この例では、被参照概念階層が赤線に囲まれた部分となる。そのため、“F1 競技”で参照されている“F1 マシン”の兄弟概念が比較対象から除外されるため、 $X=Y=0$ であり、比較対象の 3 つの is-a 階層が相似となり、概念及びスロットの追加提案ができなくなる。

また、親概念から等距離の兄弟概念に限定する場合は、“F1 マシン”は、距離が 2 であり、被参照概念階層に距離が 2 の概念がこの被参照概念階層には無い。この場合も概念の追加提案が不可能となる。

しかし、これらの提案不可能となった例であっても同じ親概念を持つ兄弟概念に区切るのではなく、兄弟概念が少なくとも一つは出てくる範囲まで拡張した場合、有効な洗練提案が可能となった。図 6 の例では、“F1 マシン”から“レース車両”まで上位を辿ると、“ツーリングカー”という兄弟概念を一つ見つけることが可能であり、“ツーリングカー”を使用するような“F1 競技”の兄弟概念を基本概念階層に追加するといった提案が可能となる。

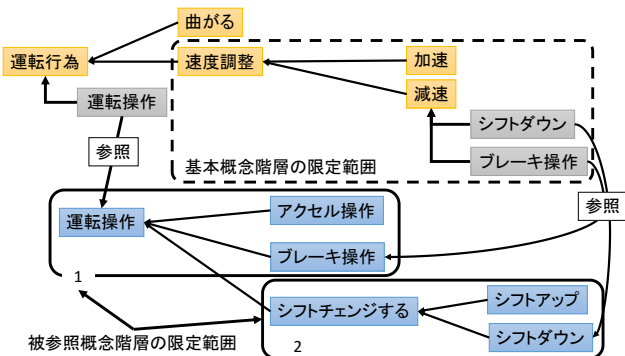


図 5 比較される概念階層の制限例

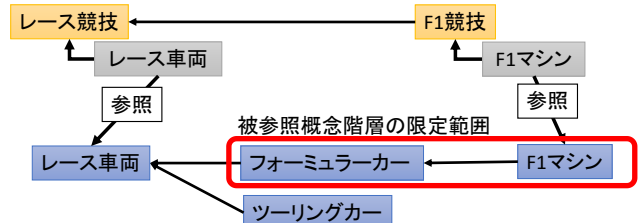


図 6 兄弟概念の限定により提案不可となった例

5. まとめ

本研究では、概念分類の一貫性の高いオントロジーは参照関係にある is-a 階層が互いに相似となるという性質に着目し、3 つの is-a 階層を比較することで洗練箇所及び方法を提案する洗練支援システムを開発した。以前までの上下の概念階層の比較に加え、今回は、兄弟概念を含む概念階層を比較する手法を開発した。さらに、比較範囲の拡大に伴う洗練提案数の増加に対応するための概念階層の範囲の限定方法の考察を行い、評価実験を行った。その結果、兄弟概念階層の比較により以前手法よりも提案の正確性を上げながら、兄弟概念階層の制限により同程度の洗練提案数できることが分かった。今後、比較対象の限定によって洗練提案不可能となったものについての考察を進めていくとともに、以前までの手法との統合を進めていきたい。

謝辞

本研究の一部は科学研究費補助金基盤研究(B)25280081 の助成による。

参考文献

[Decker 98] Decker, S., Erdmann, M., Fensel, D., and Studer, R.: Ontobroker: Ontology based Access to Distributed and Semi-Structured Information, in Database Semantics: Semantic Issues in Multimedia Systems, pp. 351-369, Kluwer Academic Publisher (1998)

[Guarino 04] Guarino N. and Welty C. An overview of OntoClean. Handbook on ontologies (pp. 151-159). Berlin; Heidelberg: Springer. 2004

[Kopena 03] Kopena, J. B. and Regli, W. C.: DAMLJessKB: A Tool for Reasoning with the Semantic Web, IEEE Intelligent Systems, Vol. 18, pp. 74-77 (2003)

[M. Ohta 11] M. Ohta, K. K. and Mizoguchi, R.: A Quality Assurance Framework for Ontology Construction and Refinement, in 7th Atlantic Web Intelligence Conference (AWIC2011) (2011)

[Sirin 07] Sirin, Evren and Parsia, Bijan and Grau, Bernardo Cuenca and Kalyanpur, Aditya and Katz, Yarden. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics, 5(2):51-53, 2007.

[太田 10] 太田衛, 古崎晃司, 溝口理一郎: オントロジーの学習の現状と動向, 人工知能学会論文誌, Vol. 126, pp. 403-418 (2010)

[増田 15] 増田壮志, 古崎晃司, 駒谷和範: 参照関係にある is-a 階層間の相似性を利用したオントロジー一貫性向上手法の評価, 人工知能学会第 29 回全国大会 (JSAI2015), 2M1-3, 2015.

[溝口 06] 溝口理一郎: オントロジー構築入門, オーム社(2006)