

オントロジーエディタ DODDLE ファミリーの開発と利用

Development and the use of Ontology Editor DODDLE Family

森田 武史*1
Takeshi Morita

山口 高平*2
Takahira Yamaguchi

*1 慶應義塾大学 大学院 理工学研究科
Graduate School of Science and Technology, Keio University

*2 慶應義塾大学 理工学部
Faculty of Science and Technology, Keio University

あらまし

マニュアル主導のみでのオントロジー構築には限界があることから、既存情報資源を活用して領域オントロジーを半自動で構築可能なオントロジー構築支援環境 DODDLE を開発してきた。また DODDLE を基礎として、セマンティック Web 対応を含めた、さまざまなバージョンの DODDLE の開発を現在進めている。本稿ではそれらを DODDLE ファミリーと呼び、DODDLE ファミリー開発の経緯、現状、利用、今後の展望について述べる。

1. はじめに

近年、セマンティック Web 技術の標準化やそれに関連する研究・技術開発が活発に行われている。オントロジーはセマンティック Web の中核技術の一つであり、計算機が処理可能なメタデータの意味情報の定義や知識を共有するために必要となる。オントロジー構築に関するさまざまな研究が行われており、セマンティック Web のためのオントロジー記述言語 OWL の標準化に伴い、オントロジー構築支援ツールが数多く開発されている [古崎 05]。オントロジーをマニュアル主導で構築するためのツールとして Protégé [Knublauch 04] や法造 [古崎 02] がある。それに対して、マニュアル主導のみでのオントロジー構築には限界があることから、既存情報資源を活用して領域オントロジーを半自動で構築可能なオントロジー構築支援環境として DODDLE (a Domain Ontology rapiD DeveLopment Environment) [山口 99] を開発してきた。DODDLE は当初、知識工学におけるオントロジー構築支援を目的として、領域知識を再利用するために、特定の知識システムのために研究開発が行われた。しかし上記の背景より、近年ではセマンティック Web におけるオントロジー構築支援に移行してきている。DODDLE のポリシーである既存情報資源の活用を基礎として、さまざまなバージョンの DODDLE の開発を現在進めている。本稿では、それらを DODDLE ファミリーと呼び、DODDLE ファミリー開発の経緯、現状、利用、今後の展望について述べる。

本稿の構成は次のようになっている。2章では DODDLE ファミリーのこれまでの経緯および現状について述べる。3章では、DODDLE ファミリーの共通アーキテクチャについて述べる。4章では、DODDLE ファミリーの利用例として、ロケット運用オントロジー及び分析クラス図作成支援について述べる。5章では本稿をまとめ、DODDLE ファミリーの今後の展望について述べる。

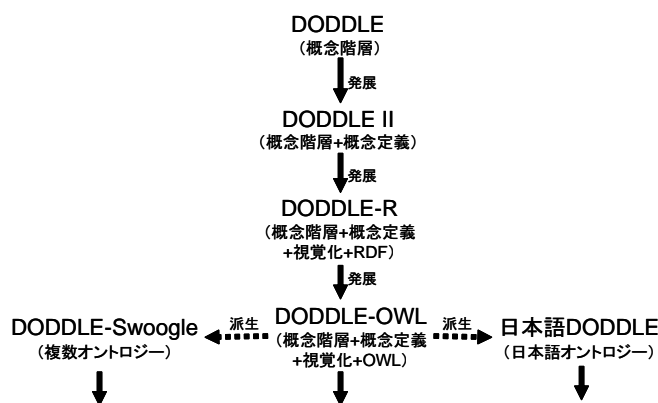


図 1: DODDLE ファミリーの系統図

2. DODDLE ファミリー

DODDLE は、既存の情報資源を活用して領域オントロジーを半自動で構築可能なオントロジー構築支援環境であり、これまでに様々なバージョンが開発されてきた。本稿では、DODDLE を基に開発されてきた様々なバージョンの DODDLE を DODDLE ファミリーと呼ぶ。図 1 に DODDLE ファミリーの系統図を示す。現在、DODDLE-OWL と DODDLE-OWL から派生した日本語 DODDLE 及び DODDLE-Swoogle の開発が行われている。DODDLE-OWL は、英語を概念の表記としてもつ領域オントロジーの構築を支援する。日本語 DODDLE は、日本語を概念の表記としてもつ領域オントロジーの構築を支援する。DODDLE-Swoogle は、オントロジー検索エンジン Swoogle を利用して、既存の複数オントロジーから領域オントロジーの構築を支援する。以下では、DODDLE-OWL、日本語 DODDLE、DODDLE-Swoogle の概要を述べる。

2.1 DODDLE-OWL

先行研究において、電子化辞書および対象領域に関する英語専門文書を用いて領域オントロジーを半自動で構築する DODDLE、DODDLE II [Kurematsu 04]、DODDLE-

連絡先: 〒 223-8522 横浜市港北区日吉 3-14-1 慶應義塾
大学理工学部管理工学科 矢上キャンパス 24 号棟 6
階 619A 号室 TEL:045-566-1614 E-mail: {t.morita,
yamaguti}@ae.keio.ac.jp

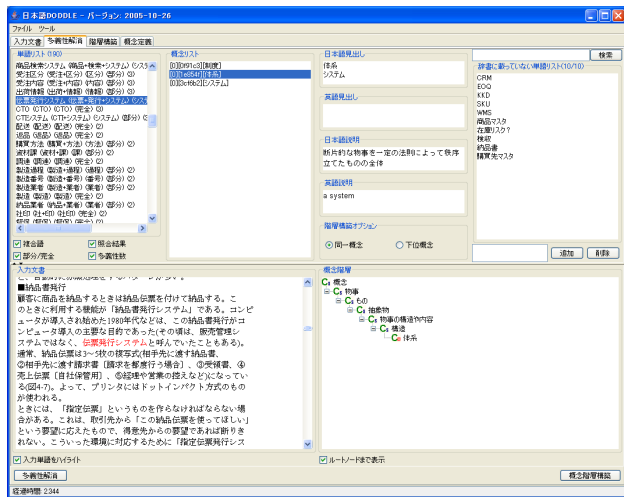


図 2: 日本語 DODDLE を用いて多義性解消を行っている際のスクリーンショット

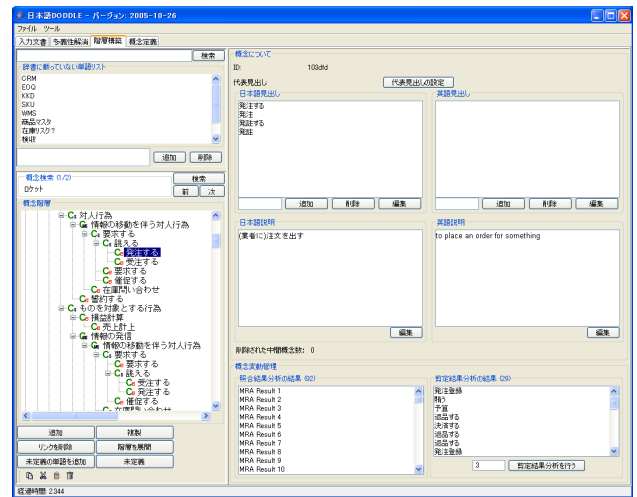


図 3: 日本語 DODDLE を用いて概念階層を構築している際のスクリーンショット

R(RDF extension) [Sugiura 04], DODDLE-OWL(OWL extension) [Morita 04b] が開発されてきた。DODDLE, DODDLE II, DODDLE-R, DODDLE-OWL では、電子化辞書として WordNet [Miller 95] を利用しており、英語を概念の表記としてもつ領域オントロジーの構築を支援している。DODDLE は、領域オントロジーにおける概念階層の構築を支援している。また、電子化辞書と対象領域における概念階層構造には、対象領域の違いにより種々の差異がある。特定の部分の概念が領域によって変化することは概念変動と呼ばれる。DODDLE は、照合結果分析と剪定結果分析の 2 つの戦略を用いて概念変動の管理を支援している。DODDLE II は、DODDLE に加えて、領域オントロジーにおける概念定義の構築を支援している。概念定義には、WordSpace 法 [Hearst 96] および相関ルール [Agrawal 94] の 2 つの共起性に基づく統計的な手法を用いている。DODDLE-R 及び DODDLE-OWL は、DODDLE II に加えて、視覚化機能およびエクスポート機能を支援している。視覚化機能では、概念階層および概念定義の編集および概念変動の視覚的な示唆、編集を支援している。エクスポート機能では、セマンティック Web におけるオントロジー記述言語 OWL(-Lite) をエクスポート可能としている。

2.2 日本語 DODDLE

DODDLE-OWL までは、英語を概念の表記としてもつ領域オントロジー構築を支援してきた。一方、日本語専門文書を利用した、日本語を概念の表記としてもつ領域オントロジー構築に関する要望も増えてきていることから、DODDLE-OWL のアーキテクチャを基に、日本語 DODDLE を新たに開発している。図 2 は、日本語 DODDLE を用いて多義性解消を行っている際のスクリーンショットである。図 3 は、日本語 DODDLE を用いて概念階層を構築している際のスクリーンショットである。オントロジーは言語に依存しないが、オントロジーを半自動構築するためには、参照リソースとして特定の言語を表記としてもつ一般的なオントロジーや対象領域に関する専門文書が必要となる。日本語 DODDLE は、EDR 電子化辞書 [日本 01] と対象領域に関する日本語専門文書を用いて、日本語を概念の表記としてもつ領域オントロジーを半自

動で構築する。EDR 電子化辞書は、単語辞書、対訳辞書、概念辞書、共起辞書、専門用語辞書（情報処理）、EDR コーパスから構成される。日本語 DODDLE では、これらの中から、単語辞書、概念辞書、専門用語辞書（情報処理）を利用している。単語辞書には、日本語単語辞書（27万語）および英語単語辞書（19万語）がある。概念辞書には、41万概念から構成される概念体系辞書および概念記述辞書がある。専門用語辞書は、日本語専門用語単語辞書（情報処理）（11万語）、英語専門用語単語辞書（情報処理）（7万語）を含んでいる。現在、概念辞書と専門用語辞書の 2 種類の辞書を切り替え可能な仕組みとなっている。また、EDR 電子化辞書は日本語と英語両方を概念の見出しとして持つため、日本語 DODDLE では日本語だけでなく、英語を概念の表記としてもつ領域オントロジーの構築も可能となっている。日本語 DODDLE は現在開発中のため、DODDLE-OWL に実装されている WordSpace 法および相関ルールを用いた概念定義の機能は実装されていない。代わりに日本語 DODDLE では、EDR 電子化辞書の概念記述辞書を利用して概念定義を行うことが可能である。EDR 電子化辞書の概念辞書には、2 つの概念の間に上位-下位以外の関係があるとき、その関係を記述した概念記述辞書がある。概念記述辞書には動詞的概念が名詞的概念を支配する場合の格関係を中心に、object, agent, goal, implement, a-object, place, scene, cause の 8 種類の概念関係が定義されている。日本語 DODDLE では、8 種類の概念関係子のうち agent および object を利用している。動詞的概念をオブジェクトプロパティ、動詞的概念と agent 関係がある名詞的概念をプロパティの定義域、動詞的概念と object 関係にある名詞的概念をプロパティの値域として定義している。

2.3 DODDLE-Swoogle

DODDLE-OWL 及び日本語 DODDLE では、WordNet または EDR 電子化辞書という単一の汎用オントロジーを参照リソースとして領域オントロジーにおける概念階層および概念定義の構築を試みてきた。近年、セマンティック Web 技術の標準化やそれに関連してオントロジー構築支援ツールも多数開発されてきており [古崎 05]、領域オントロジーの数も増えつ

つある。OWL では、多数の Web 上に分散して開発されたオントロジーをインポート、マッピングしながら、オントロジーを構築するための仕組みが用意されている。単一オントロジーの構築支援だけでなく、複数オントロジーから必要な部分を抽出、合成して、領域オントロジーの構築を支援する必要がある。DODDLE ファミリーの既存情報リソースを再利用するというポリシーから、既存の Web 上に存在する複数の領域オントロジーを再利用して、領域オントロジーの構築支援を行う DODDLE-Swoogle を現在開発中である。

既存の領域オントロジーを再利用するためには、ユーザが構築を試みている対象領域に関連する領域オントロジーを検索する必要がある。Web 上に存在する領域オントロジーを検索するための方法としては、検索エンジン google を利用することが可能だが、オントロジー以外の Web 文書も検索されてしまうため、適切なオントロジーを検索することは困難である。さらに、概念階層の構築を支援するためには、クラス単位の検索を行いたいこともあり、オントロジーの検索に特化した検索エンジンが必要となる。既存のオントロジー検索エンジンとしては、Swoogle [Ding 05] や Ontaria [W3C] がある。DODDLE-Swoogle では、Swoogle をオントロジー検索に用いている。Swoogle では、クラス単位、プロパティ単位の検索やオントロジー内に明示的に記述されてない逆リンクの関係（あるクラスを参照しているインスタス一覧など）を検索することが可能である。また、クラス及びプロパティを検索するための Web Service (Swoogle Web Service) も提供されており、プログラム上からオントロジーを検索することも可能である。Swoogle には、google のページランクと同様に、オントロジーのためのページランク (OntoRank) の仕組みが用意されており、同一ローカル名をもつクラス及びプロパティに対してランキングされている。OntoRank では、より多くのオントロジーから参照されているクラスまたはプロパティを高くランキングしているが、領域オントロジー構築の観点からは、必ずしも多数のオントロジーから参照されているオントロジーが利用可能とは限らない。ユーザが構築の対象としている領域に関連する領域オントロジーを適切に検索するための仕組みが必要であるが、現在検討中である。

現状では 3 章で述べる DODDLE ファミリーの共通アーキテクチャを基に、対象領域に関連する重要語句からクラスを検索し、対象クラスからルートクラスまでのパスを抽出する部分までの実装が完了している。上位オントロジーの構造の異なる複数のオントロジーをどのように合成していくかは今後の課題である。また、DODDLE で支援している概念変動の管理も考慮していく予定である。

3. DODDLE ファミリーの共通アーキテクチャ

図 4 に DODDLE ファミリーの共通アーキテクチャを示す。DODDLE ファミリーは、主に、入力モジュール、オントロジー構築モジュール、オントロジー洗練モジュール、視覚化モジュール、変換モジュールの 5 つのモジュールから構成される。DODDLE ファミリーでは、1 つ以上の領域に関連する専門文書があること及び領域における重要な語彙をユーザが選択できることを前提とする。

はじめに、ユーザは入力モジュールにおいて、領域において重要な語句と電子化辞書中の概念との対応付けを行う（入力概

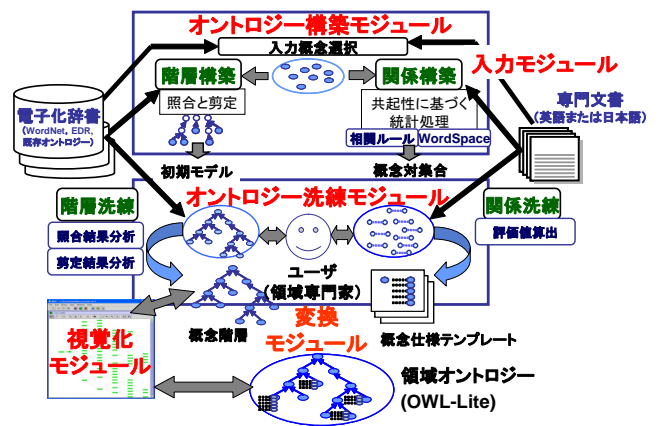


図 4: DODDLE ファミリーの共通アーキテクチャ

念の選択)。DODDLE-OWL 及び日本語 DODDLE では、多義性をもつ語句については、全入力単語と電子化辞書概念階層を利用してランキングをして、優先度の高い意味から順番に概念を表示するようにしている。DODDLE-OWL では、入力モジュールにおいて WordNet を利用して名詞と動詞を抽出し、原型に変換し、出現頻度順にソートして出力する。WordNet にアクセスするための API としては、JWNL (Java WordNet Library) [Didion] を利用している。日本語 DODDLE では、Java で実装された形態素解析器 Sen [Sen] を利用して文書中の名詞、動詞、その他の品詞を抽出可能としている。また、専門用語自動抽出システム [中川 03] を用いて複合語の抽出も可能としている。大部分の複合語については、電子化辞書中の概念と完全に対応付けすることが困難なため、複合語を Sen を用いて形態素解析を行い、先頭の単語を順に除いて電子化辞書中の概念と対応付けを試していき、最長一致した概念と対応付けを行う。例えば、「優良顧客」が電子化辞書中の概念の表記として存在しない場合には、「優良」と「顧客」に形態素解析し、「顧客」で電子化辞書中の概念との対応付けを試みて、一致した場合には、「優良顧客」と電子化辞書中の「顧客」概念の対応付けを行う。その際に、「顧客」概念の下位概念とするか、「顧客」概念の別の表記として「優良顧客」を用いるかをユーザは選択可能である。

オントロジー構築モジュールは、オントロジーの基礎となる初期概念階層と概念対のセットを電子化辞書と専門文書を参照しながら、入力概念を基に生成する。DODDLE-OWL では、英語専門文書と電子化辞書として WordNet を参照する。日本語 DODDLE では、日本語専門文書と電子化辞書として EDR 電子化辞書を参照する。DODDLE-Swoogle では、電子化辞書として既存の領域オントロジーを Swoogle を利用して検索し、参照する。初期概念階層は IS-A 階層として構築される。概念対のセットは共起性に基づく統計処理を用いて獲得される。これらの概念対の中から特に重要な概念対を選択し、その間の関係を定義したものが非階層関係 (概念定義) となる。

オントロジー洗練モジュールでは、ユーザとのインタラクションを通してオントロジー構築モジュールで構築された初期オントロジーを洗練していく。初期オントロジーを洗練するために、概念変動と概念対のセットの評価の管理を支援する。初期概念階層は一般的なオントロジーから生成されるため、ユー

は概念変動と呼ばれる問題を考慮しながら、初期概念階層を特定の領域に調整する必要がある。それは、特定の部分の概念が領域によって変化することを意味する。概念変動管理のために、DODDLE ファミリーは照合結果分析と剪定結果分析の2つの戦略を適用する。これらの戦略の詳細は [山口 99] で詳述されている。オントロジー構築モジュールで生成された概念対のセットから概念仕様テンプレートを構築するためには、重要概念対を評価するための指標が必要となる。[Kurematsu 04] では、WordSpace 法 [Hearst 96] による文脈類似度と相関ルール [Agrawal 94] による信頼度の2つの共起性に基づく統計的な手法を用いた重要概念対の評価方法が提案されている。これらの手法は、専門文書中における重要概念の共起に基づいており、特定の領域に依存することなく利用することができる。

DODDLE-OWL 及び日本語 DODDLE には、領域オントロジーの質を高めるために、視覚化モジュールが導入されている。視覚化モジュールには MR³ (Meta-Model Management based on RDFs Revision Reflection) [Morita 04a] を利用している。MR³ とは RDF と RDFS・OWL(-Lite) をモデルとメタモデルの関係としてとらえ、両者の視覚的編集および一貫性を半自動的に管理するツールである。MR³ のプラグイン機構を用いて、DODDLE-OWL または日本語 DODDLE と相互に OWL データの交換を行う。領域オントロジー構築における視覚化モジュールの役割は二つある。一つ目はオントロジー洗練モジュールにおける概念変動管理を視覚的に支援する機能である。視覚化モジュールにオントロジー構築モジュールで構築された初期概念階層を表示し、ユーザは DODDLE ファミリーが示唆する概念変動が生じる可能性のある箇所の編集を行う。二つ目はオントロジーの外在化である。オントロジーの外在化とは概念階層と概念定義を DODDLE ファミリー以外の見方によって視覚的に表示することを意味する。オントロジーの外在化を行うことによって、オントロジー全体 (概念階層及び概念定義) のバランスを見ながら調整を行い、オントロジーの質を向上させる。MR³ のグラフィブラリの制限から数百概念程度までしか概念階層の視覚化、編集を行うことができないため、日本語 DODDLE では Java-Swing の JTree ライブラリを利用して、独自の概念階層エディタを実装しており、数万概念程度の概念階層の視覚化及び編集を行うことが可能となっている。

最終的に構築されたオントロジーは、変換モジュールによって OWL 形式に変換される。OWL 形式の変換には Jena セマンティック Web フレームワーク [HP Labs] を利用している。

実装については、DODDLE は Tcl/Tk 及び Perl、DODDLE II は Perl 及び Perl/Tk、DODDLE-R、DODDLE-OWL、日本語 DODDLE、DODDLE-Swoogle は Java 言語で実装されている。Java 言語で実装する理由の一つは、Jena などのセマンティック Web 技術を扱うためのライブラリが Java 言語で多数開発されているためである。

4. ケーススタディ

本章では、DODDLE ファミリーの利用について述べる。これまでに、DODDLE ファミリーは法律オントロジー [山口 99, Kurematsu 04]、ビジネスオントロジー [Kurematsu 04]、ロケット運用オントロジー [洪 05]、分析クラス図作成支援における領域オントロジーの構築に利用されている。ロケット運用オントロジーと分析クラス図作成支援における領域オントロジー

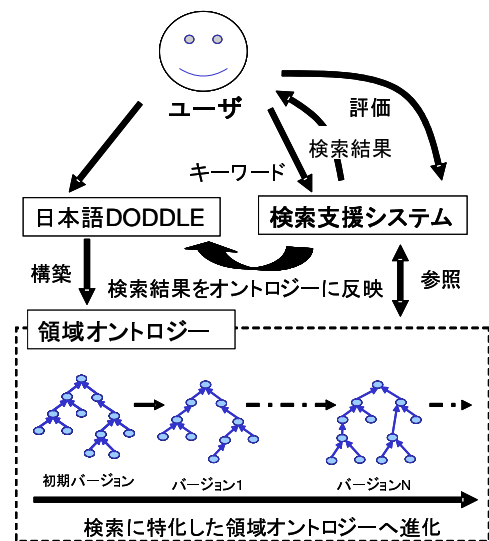


図 5: オントロジーを用いた検索システムのシステムフロー

の評価は現在行っているところである。以下では、DODDLE ファミリー (特に日本語 DODDLE) がどのような局面で利用されているか、また、どのような手順で領域オントロジーの構築を行っているかを中心に述べる。

4.1 ロケット運用オントロジーの構築

より精度の高い検索を実現するために、従来の検索システムのような検索キーワードによる全文検索ではなく、文書の意味を考慮した意味検索に関する研究が多数行われている。オントロジーを用いて意味検索をするための研究も行われているが、オントロジーは構築コストが高くボトルネックとなっている。[洪 05] は、オントロジーを用いた意味検索および検索結果に対するユーザからのフィードバックを得ることによって、オントロジーを洗練しながら検索に特化したオントロジーを構築する提案を行っている。[洪 05] では、最初に検索システムに搭載するオントロジー (初期オントロジー) の構築コストを削減するために、日本語 DODDLE を用いて初期オントロジーを半自動構築する方法を用いている。図 5 に日本語 DODDLE を初期オントロジー構築に用いた検索システムのシステムフローを示す。

以下では、[洪 05] の評価実験で行われたロケット運用オントロジーの構築について述べる。評価実験は、(株) Galaxy Express (GE) 社の協力によって、ロケット運用に関する 10,500 個の文書を対象に行った。評価実験では、ロケット運用に関する文書からロケット運用オントロジーを日本語 DODDLE を用いて構築し、検索実験を行っているが、ここではロケット運用オントロジーの構築についてのみ述べる。

はじめに、形態素解析システム茶筌 [松本 00] により、文書からキーワードを抽出する。前処理として、形態素解析を適切に行うために半角文字を全角文字に変換する。文書から抽出された一般名詞、その他の名詞、動詞、未知語の数を表 1 に示す。

評価実験では、抽出されたキーワードの中から、一般名詞、その他の名詞、動詞の計 13,541 語を日本語 DODDLE の入力とした。その際に、1 文字の単語および重複する単語は除去し

表 1: キーワード抽出結果

| 品詞名 | 語数 |
|--------|----------|
| 一般名詞 | 6,926 語 |
| その他の名詞 | 6,192 語 |
| 動詞 | 3,617 語 |
| 未知語 | 6,211 語 |
| 合計 | 22,945 語 |

ている。次に、日本語 DODDLE を用いて EDR 電子化辞書に含まれている概念の見出し（表記）と入力した単語とのマッチングを行った。10,301 語が EDR に含まれており、その中で多義性をもつ約 6,000 語について、手動で多義性解消を行った。多義性解消は 2 人で分担して行い、約 15 時間かかった。次に、概念階層を自動構築し、10,177 概念の階層を構築した。本来は最後に概念階層の洗練および概念変動の管理を行うために、DODDLE で提案された 2 つの戦略（照合結果分析および剪定結果分析）を適用すべきであるが、本実験では概念階層の概念数が 10,000 概念以上と膨大な数であったために、照合結果分析が示唆する部分木が数千箇所にわたり、ユーザ手動で結果を確認していくことは困難であった。

4.2 分析クラス図作成支援

オブジェクト指向に基づくソフトウェア開発において、分析クラス図の構築は上流工程の要の一つであるが、現状はモデラーのスキル依存で品質にばらつきが大きい。[神谷 05]では、オントロジーに基づいてデータ構造化された業務概念体系を用いて、要求文書等から分析クラス図案を導出することでモデラーを支援する方法を提案している。図 6 に、分析クラス図作成支援体系を示す。分析クラス図作成支援体系は、領域オントロジー構築支援系と分析クラス図作成支援系の 2 つから構成される。前者では業務知識関連文書を入力として、汎用オントロジーを用いて業務専門家の支援下で領域オントロジーの構築を行う。後者では要求仕様関連文書を入力として、領域オントロジー内の情報を活用しながら、モデラーが個別情報システム開発プロジェクトの分析クラス図を作成する。日本語 DODDLE は、図 6 の領域オントロジー構築支援系として利用されている。

現在、コンピュータの販売管理に関する要求仕様からの分析クラス図作成を、図 6 の分析クラス図作成支援体系に基づいて行っている。物流・在庫・販売管理に関する領域オントロジーの構築を試みている。領域オントロジーの構築手順を図 7 に示す。はじめに業務知識関連文書として約 26,000 語の物流・在庫・販売管理に関する文書から形態素解析及び専門用語抽出システムを用いて、名詞および動詞 1,620 語を抽出する。次に抽出された語句の中から領域にとって重要な語句をユーザが手作業で選択する。その後、日本語 DODDLE に領域にとって重要な語句を入力し、初期領域オントロジーを生成する。最後に、ユーザが概念変動箇所の修正や EDR 電子化辞書に載っていない語句を概念階層に追加し、領域オントロジーが完成する。

4.3 考察

ロケット運用オントロジー構築時の問題点は、10,500 個という大量の文書からオントロジーを構築するために入力語彙

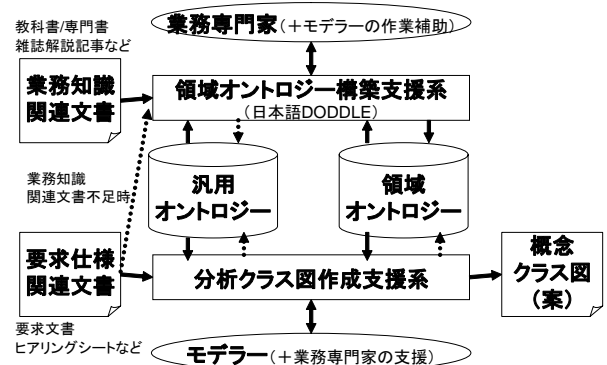


図 6: 分析クラス図作成支援体系

1. 物流・在庫・販売管理に関する文書（日本語, 30ページ, 約26,000語）

↓ 形態素解析, 専門用語抽出システム

2. 単語抽出（名詞と動詞, 1,620単語）

↓ 手作業1

3. 抽出単語の洗練・分類（202単語）

↓ 日本語DODDLE

4. 初期領域オントロジー作成

↓ 手作業2

5. 領域オントロジー作成（226概念）

図 7: 領域オントロジー構築手順

数が膨大となり、前処理である多義性解消にかなりの時間がかかってしまうことであった。DODDLE には自動的に多義性を解消するための仕組みが用意されており、法律オントロジー構築時には、入力語彙数が約 50 と少数であったために 8 割から 9 割の精度で多義性解消ができた。今回は入力語彙数が約 10,000 と膨大であったために、手動で多義性解消した結果と自動で多義性解消した結果を比較した場合、4 割から 5 割の精度であった。すべての入力語彙集合を利用するのではなく、文書単位に分割して多義性解消を行えば精度は向上すると考えられる。概念変動管理についても、概念数が膨大であったために、照合結果分析については概念変動が生じる可能性がある箇所の提示が数千個と膨大な数になり、ユーザがすべてを確認することは困難であった。概念数が膨大であるために、本質的に概念変動が生じる箇所も多いとも考えられるが、新たな概念変動を管理するための戦略を検討する必要があると思われる。また、EDR 電子化辞書と入力語彙の照合を行う際に、専門性の高い語句が数多く照合されなかった。特にロケット運用に関する文書の中には、専門用語の他に、専門用語の省略語や企業固有の用語が含まれているため、そのような語句については、汎用オントロジーである EDR 電子化辞書では照合できなかった。省略語については展開することができれば、末尾の語句と

EDR 電子化辞書との照合は可能だと考えられる。汎用オントロジーと照合できなかった語句を概念階層中のどこに位置づけるか示唆するための機構については、今後の課題である。

分析クラス図作成支援における領域オントロジーの構築では、クラスと属性の区別が重要となる。現状では、図7の手作業1においてユーザが手動で洗練・分類を行っているが、EDR 電子化辞書の上位概念を用いて半自動的に分類が可能かどうか現在検討している。例えば、EDR 電子化辞書には「事物の属性」という意味の概念が存在するが、その下位概念に位置する概念は属性とみなすことができると考えられる。また、分析クラス図を構築するためには、クラス間の関連の同定が重要であるが、それを支援するためには、領域オントロジーに IS-A 以外の関係を定義した概念定義が必要となる。概念定義は DODDLE ファミリーでは、プロパティとその定義域および値域として表現している。現在は、定義域および値域は全てクラスとして、プロパティは OWL でいうところのオブジェクトプロパティとして扱っている。分析クラス図作成のための領域オントロジーとしては、属性と関連を区別するために、値域としてデータ型をとるデータタイププロパティを扱えるようにする必要があると考えられる。

5. おわりに

本稿では、オントロジーエディタ DODDLE ファミリー開発の経緯、現状、共通アーキテクチャについて述べた。また、ロケット運用オントロジーおよび分析クラス図作成支援に DODDLE ファミリー（日本語 DODDLE）がどのように利用されているかを述べた。考察では DODDLE ファミリーを用いた領域オントロジー構築支援における今後の課題を述べた。今後はユーザとのインタラクションを重視しながら、現在手動で行っている作業をできるだけ半自動で処理できるようにしていく予定である。DODDLE-Swoogle により複数の既存領域オントロジーを再利用して、領域オントロジーを半自動構築できるようになれば、DODDLE-OWL および日本語 DODDLE を統合することも可能だと思われる。DODDLE-OWL については [DODDLE-Project] よりダウンロード可能となっている。ユーザからのフィードバックを参考に、セマンティック Web におけるオントロジー構築支援手法の改良について考察を行っていくことも今後の課題である。

参考文献

[Agrawal 94] Agrawal, R. and Srikant, R.: Fast algorithms for mining association rules, pp. 487–499 (1994)

[Didion] Didion, J.: JWNL(Java WordNet Library): <http://sourceforge.net/projects/jwordnet>

[Ding 05] Ding, L., Pan, R., Finin, T., Joshi, A., Peng, Y., and Kolari, P.: Finding and Ranking Knowledge on the Semantic Web, *Proceedings of the 4th International Semantic Web Conference, LNCS 3729*, pp. 156–170 (2005), <http://swoogle.umbc.edu/>

[DODDLE-Project] DODDLE-Project, : <http://mmm.semanticweb.org/doddle/>

[Hearst 96] Hearst, M. A. and Schutze, H.: Customizing a Lexicon to Better Suit a Computational Task, *Corpus Processing for Lexical Acquisition*, pp. 77–96 (1996)

[HP Labs] HP Labs, : Jena Semantic Web Framework: <http://jena.sourceforge.net/downloads.html>

[Knublauch 04] Knublauch, H., Fergerson, R. W., Noy, N. F., and Musen, M. A.: The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications, *Third International Semantic Web Conference*, pp. 229–243 (2004)

[Kurematsu 04] Kurematsu, M., Iwade, T., Nakayama, N., and Yamaguchi, T.: DODDLE II A Domain Ontology Development Environment Using a MRD and Text Corpus, *IEICE(E)*, Vol. E87-D, No.4, pp. 908–916 (2004)

[Miller 95] Miller, G. A.: WordNet: A Lexical Database for English, Vol. 38, No.11, pp. 39–41 (1995)

[Morita 04a] Morita, T., Izumi, N., Fukuta, N., and Yamaguchi, T.: MR^3 : Meta-Model Management based on RDFs Revision Reflection, *Proceedings of the 6th Joint Conference on Knowledge-Based Software Engineering (JCKBSE)*, pp. 228–236 (2004)

[Morita 04b] Morita, T., Shigeta, Y., Sugiura, N., Fukuta, N., Izumi, N., and Yamaguchi, T.: DODDLE-OWL: OWL-based Semi-Automatic Ontology Development Environment, *Evaluation of Ontology-based Tools* (2004), <http://mmm.semanticweb.org/doddle/>

[Sen] Sen Project, <http://ultima.org/sen/>

[Sugiura 04] Sugiura, N., Shigeta, Y., Fukuta, N., Izumi, N., and Yamaguchi, T.: Towards On-the-fly Ontology Construction - Focusing on Ontology Quality Improvement, *1st European Semantic Web Symposium* (2004)

[W3C] W3C, : Ontaria: <http://www.w3.org/2004/ontaria/>

[古崎 02] 古崎 晃司, 來村 徳信, 池田 満, 溝口 理一郎: 「ロール」および「関係」に関する基礎的考察に基づくオントロジー記述環境の開発, *人工知能学会論文誌*, Vol. 17, No.3, pp. 196–208 (2002)

[古崎 05] 古崎 晃司, 溝口 理一郎: オントロジー構築ツールの現状, *人工知能学会誌*, Vol. 20, No6, pp. 707–714 (2005)

[洪 05] 洪 潤基, 森田 武史, 飯島 正, 小出 誠二, 山口 高平: 日本語 DODDLE に基づくロケット運用オントロジーの構築と評価, *信学技報 (知能ソフトウェア工学研究会)* (2005)

[山口 99] 山口 高平, 樽松 理樹, 青木 千鶴, 関内 律恵子, 加賀 山茂, 吉野 一: 計算機可読型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, pp. 1080–1087 (1999)

[松本 00] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原 正幸: 日本語形態素解析システム『茶筌』 version 2.2.1 使用説明書 (2000), <http://chasen.naist.jp/hiki/ChaSen/>

[神谷 05] 神谷慎吾, 石橋昌彦, 森田武史, 福田直樹, 飯島正, 山口高平: オントロジーを用いた分析クラス図作成支援, *信学技報 (知能ソフトウェア工学研究会)*, Vol. 105, No.208, pp. 25–33 (2005)

[中川 03] 中川裕志, 森辰則, 湯本紘彰: 出現頻度と接続頻度に基づく専門用語抽出, *自然言語処理*, Vol. 10, No. 1, pp. 27–45 (2003), <http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>

[日本 01] 日本電子化辞書研究所: EDR 電子化辞書 (第2版) 仕様説明書 TR-006 (改) (2001)