

オントロジー検索エンジンを用いた 領域オントロジー構築支援環境 DODDLE-OWL の拡張

Extensions of a Domain Ontology Development Environment: DODDLE-OWL Using an Ontology Search Engine

森田 武史*¹ 和泉 憲明*² 山口 高平*¹
Takeshi Morita Noriaki Izumi Takahira Yamaguchi

*¹慶應義塾大学
Keio University

*²産業技術総合研究所
Advanced Industrial Science and Technology

あらまし 本研究では、オントロジー検索エンジン Swoogle を用いて領域オントロジー構築支援環境 DODDLE-OWL の拡張を行う。また、Swoogle を用いて獲得した OWL オントロジー及び専門文書を参照リソースとした、領域オントロジー構築支援手法について述べる。

1. はじめに

近年、Web コンテンツに計算機が理解可能なメタデータを付与することにより、現在よりも知的な情報探索や動的な Web サービス連携を可能にするセマンティック Web に関連する研究が活発に行われている。セマンティック Web を実現するためには、メタデータ(概念)の意味を定義するオントロジーを整備する必要がある。しかし、特定の領域概念を扱った領域オントロジーは、問題領域に含まれる概念数が膨大であること、領域概念の意味の専門性の高さなどから、多大な開発コストを要するという問題がある。

先行研究において、日本語及び英語概念を対象とした領域オントロジー構築支援環境である、DODDLE-OWL [Morita 06] 及び日本語 DODDLE [森田 06] を開発してきた。DODDLE-OWL 及び日本語 DODDLE では、WordNet や EDR 電子化辞書 [Yokoi 95] 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の構築および概念定義支援を試みてきた。先行研究の欠点として、既存オントロジーの再利用ができないという点があげられる。セマンティック Web におけるオントロジーは Web 上に散在するオントロジーをインポート、マッピングしながら目的に合ったオントロジーを構築することが可能であり、OWL(Web Ontology Language) にはそのための仕組みが導入されている。また、既存オントロジーの中には、特定の領域に特化したオントロジーが存在する可能性があり、ユーザの利用目的に合った既存オントロジーが存在すれば、大部分の再利用が可能になると考えられる。

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。また、Web 上に存在する OWL オントロジーを検索可能なオントロジー検索エンジンも開発されている。現状のオントロジー検索

エンジンは、特定のキーワードを含むオントロジー、クラス、プロパティなどを検索することは可能だが、領域オントロジー構築に利用可能な既存オントロジーを獲得するための仕組みはない。ユーザは既存オントロジーを再利用するために、様々なオントロジーを検索して、オントロジーの定義内容を確認しながら、目的に合ったオントロジーをマニュアルで構築する必要があり、その作業にはコストがかかる。

以上の背景より、本研究では Web 上に存在する複数の既存 OWL 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ことができるように DODDLE-OWL の拡張を行う。また、オントロジー検索エンジンを用いて獲得した OWL オントロジー及び専門文書を参照リソースとした、領域オントロジー構築支援手法について述べる。

2. オントロジー検索エンジン

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、google などの検索エンジン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、クラスやプロパティ単位での検索も必要となるため、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、Swoogle [Ding 05] や Ontaria [W3C] などがある。Swoogle には、2007 年 3 月現在 1 万以上のオントロジーが登録されている。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い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ンス一覧など)を検索することが可能である。また、オントロジーを検索するための 19 種類の REST 形式の Web サービス(Swoogle Web サービス)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。DODDLE-OWL から既存オントロジーを再利用するためには、プログラムからオントロジーを検索することが必要となるため、

連絡先: 〒 223-8522 神奈川県横浜市港北区日吉 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

Web サービスを提供している Swoogle を DODDLE-OWL で利用する。

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ランキングの仕組み (OntoRank) やクラス及びプロパティのためのランキングの仕組み (TermRank) が導入されている。OntoRank や TermRank では、より多くのセマンティック Web 文書から参照されているクラスや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、必ずしも多数のセマンティック Web 文書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。

3. オントロジー検索エンジンを用いた領域オントロジー構築支援

図 1 にオントロジー検索エンジン Swoogle を用いた領域オントロジー構築支援環境 DODDLE-OWL のシステムフローを示す。DODDLE-OWL の入力専門文書であり、出力は OWL 形式の領域オントロジーである。DODDLE-OWL により構築される領域オントロジーは、階層関係および非階層関係である。

はじめに、領域専門文書から形態素解析器を用いて名詞及び動詞を抽出する。ユーザは抽出された単語の中から領域にとって重要な単語 (入力単語) を選択する。入力単語を元に、Swoogle を用いて領域に関連する既存 OWL オントロジーの獲得およびランキングを行う。ランキングを参考にしながら、獲得された既存 OWL オントロジーの中からユーザは領域オントロジー構築に必要なオントロジー (参照オントロジー) を選択する。領域によっては既存 OWL オントロジーが存在しないことが考えられる。また、英語を概念の表記としてもつ既存 OWL オントロジーの数は増えつつあるが、日本語を概念の表記としてもつ既存 OWL オントロジーの数は現状では少ない。そのため、DODDLE-OWL では網羅性の高い汎用オントロジーを参照オントロジーとして利用することで、既存 OWL オントロジーが発見できない場合に対応する。DODDLE-OWL では参照オントロジーとして、WordNet 及び EDR 電子化辞書を利用できる。次に、参照オントロジーから入力単語を表記として持つ概念を抽出する。入力単語を表記として持つ概念が複数ある場合には、領域に最も適した概念 (入力概念) をユーザは選択する。参照オントロジー中の入力概念を含むパスを合成することにより、クラス及びプロパティ階層を構築する。Web 上に散在する異種のオントロジーのパスを合成する際には、上位概念階層の構造の違いにより単純に合成することは困難である。そのため、オントロジーアライメントによる類似概念の同定が必要となる。既存 OWL オントロジーの中から入力概念を定義域および値域とするプロパティを抽出することで、非階層関係の構築を行うことができる。また、先行研究 [Morita 06] より WordSpace 及び相関ルールを用いた非階層関係の構築支援も行うことができる。DODDLE-OWL により半自動構築された初期領域オントロジーをユーザが洗練することにより、最終的に領域オントロジーが構築できる。

Swoogle を用いて領域オントロジー構築支援を行うために DODDLE-OWL に、主に以下の 4 つの拡張を行う必要がある。

1. Swoogle を用いた既存 OWL オントロジーの獲得
2. 既存 OWL オントロジーのランキング

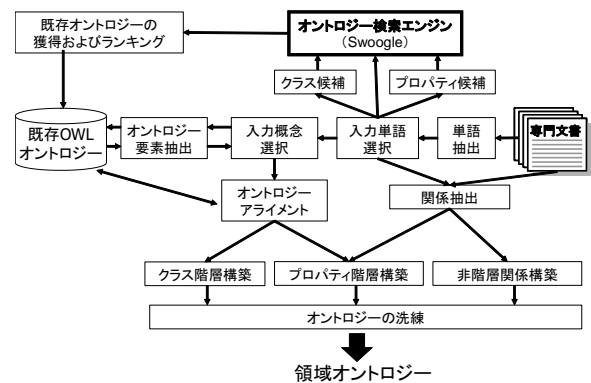


図 1: Swoogle を用いた DODDLE-OWL のシステムフロー

3. 既存 OWL オントロジーからの要素抽出
4. オントロジーアライメントを用いた階層構築

4 については現在検討中のため、以下では 1 から 3 について説明する。

4. Swoogle を用いた既存 OWL オントロジーの獲得

DODDLE-OWL は Swoogle [Ding 05] を用いて既存 OWL オントロジーの獲得を行う。Swoogle は 19 種類の REST 形式の Web サービスを提供している*1。ユーザは URL を用いてクエリーを作成し、RDF/XML 形式の検索結果を得ることができる。Swoogle が提供している Web サービスは、Search services, Digest services, Relation services の 3 つに分類できる。Search services は、検索キーワードを入力として、検索キーワードに関連する SWD (Semantic Web Document), SWO (Semantic Web Ontology), SWT (Semantic Web Term) が検索できる。SWD は RDF/XML, N-Triple, N3 形式で記述された RDF 文書を表す。SWO はクラス及びプロパティの定義の割合が 8 割以上の SWD を表す。SWT はクラスまたはプロパティを表す。Digest services は、SWD, SWT, 名前空間を入力として、それらのメタデータ (ファイルエンコーディング, RDF の形式など) を得ることができる。Relation services は、SWD, SWT を入力として、それらを参照している SWD のリストを得ることができる。また、逆に SWD, SWT が参照、依存している SWD, SWT のリストを得ることができる。表 1 に領域オントロジー構築支援に利用可能な Swoogle Web サービスとその入出力を示す。表 1 に示す Swoogle Web サービスを以下の手順で呼び出すことにより、対象領域に関連する既存 OWL オントロジーを獲得する。

1. [3] を用いて、入力単語に関連するクラス及びプロパティを獲得
2. [17] 及び [19] を用いて、1 で獲得したクラスを定義域または値域とするプロパティを獲得

*1 http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=search_overview

表 1: 領域オントロジー構築支援に利用可能な Swoogle Web サービスとその入出力

| 番号 | Swoogle Web サービス | 入力 | 出力 |
|------|-----------------------------------|---------|-------------------------------------|
| [1] | Search ontology | 検索キーワード | 検索キーワードに関連する SWO のリスト |
| [3] | Search terms | 検索キーワード | 検索キーワードに関連する SWT のリスト |
| [4] | Digest semantic web document | SWD | SWD の Swoogle メタデータ |
| [13] | list documents using term | SWT | SWT を定義, 参照, populate している SWD のリスト |
| [16] | list domain classes of a property | プロパティ | 入力したプロパティの定義域のリスト |
| [17] | list properties of a domain class | クラス | 入力したクラスを定義域とするプロパティのリスト |
| [18] | list range classes of a property | プロパティ | 入力したプロパティの値域のリスト |
| [19] | list properties of a range class | クラス | 入力したクラスを値域とするプロパティのリスト |

- [16] 及び [18] を用いて, 1 と 2 で獲得したプロパティの定義域および値域を獲得
- [1], [4], [13] を用いて, 1 から 3 で獲得したクラスまたはプロパティを定義しているオントロジーを獲得
- 1 及び 2 で獲得したプロパティの中で, 定義域および値域が入力概念または入力概念の上位概念でないものを削除
- 獲得した既存 OWL オントロジーをランキングし, ユーザが参照オントロジーを選択

Swoogle Web サービスを呼び出す際には, queryType, searchString, key の主に 3 つのパラメータを与える. queryType には, 呼び出す Web サービスを特定するためのサービスタイプを指定する. searchString には, サービスの入力となる検索キーワード, SWT, SWD などを与える. key は, 出力する検索結果数を増減させることなどができる. 初期状態では key には demo という値を指定し, Search services では, 最初の 25 個の結果のみが得られるようになっている.

1 では [3] を用いて, 入力単語をローカル名またはラベルとして持つクラス及びプロパティを獲得する. [3] では searchString パラメータに以下の検索キーワードを指定することにより, 入力単語に関連するクラス及びプロパティを獲得できる^{*2}.

- (localname:入力単語 OR label:入力単語) (type:owl.class OR type:rdfs.class)
- (localname:入力単語 OR label:入力単語) (type:owl.property OR type:rdfs.property)

また, [3] では, 入力単語に関連するクラス及びプロパティについて, Swoogle が求めた TermRank も得ることができる. TermRank は, 入力概念の候補が複数ある場合に参考となる.

1 から 3 により, クラス及びプロパティの URI を得ることはできるが, それらを定義しているオントロジーは獲得できていないため, 4 では [1], [4], [13] を用いて, 1 から 3 で獲得したクラスまたはプロパティを定義しているオントロジーを獲得する. [1] は, searchString パラメータに検索キーワー

*2 [3] で指定可能な検索キーワードについては, 以下の URL が参考になる.
http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=search_swt

ド「def:入力単語」を与えることにより, 入力単語に関連するオントロジーを獲得することができる^{*3}. また, [1] では, Swoogle が求めた OntoRank も得ることができる. OntoRank は, 対象領域に関連する既存オントロジーのランキングに用いる. [1] は, ローカル名, rdfs:label, rdfs:comment の一部に, 入力単語が含まれているオントロジーを獲得するため, [1] だけでは現状では, 1 から 3 で獲得したクラスまたはプロパティ以外のクラスやプロパティを定義しているオントロジーまで獲得してしまう. そのため, DODDLE-OWL では, [1] で得られた上位 25 個のオントロジーの中に 1 から 3 で獲得したクラスまたはプロパティが定義されていない場合には, [4] 及び [13] を用いている. [13] は, searchString にクラスまたはプロパティを与えることで, 指定したクラスまたはプロパティを定義しているオントロジーの URI を得ることができる. [4] は, [13] で得られたオントロジーの URI を searchString に与えることで, オントロジーのファイルエンコーディングと RDF の形式を得ることができる. ファイルエンコーディングと RDF の形式は, プログラムから OWL オントロジーを操作する際に必要となる. [1] の検索結果には, ファイルエンコーディングと RDF の形式が含まれているため, [4] を実行する必要はない. 1 から 3 で獲得したクラスまたはプロパティに関連するオントロジーのみを獲得したい場合には, [4] と [13] のみを実行すれば良いが, その場合にはクラスとプロパティの数だけ Web サービスを呼び出さなければいけないために, Web サービスの呼び出し回数及び処理時間がかかるという欠点がある. [1] は一度に複数の関連するオントロジーを獲得できることと, 獲得したオントロジーに対して Swoogle が求めた OntoRank が得られるため, DODDLE-OWL では, [1], [4], [13] の 3 つの Web サービスを用いている.

5 では, 4 で獲得したオントロジーを参照して, 獲得したプロパティの中で, 定義域および値域が入力概念または入力概念の上位概念でないものを削除する. これにより, 概念定義に利用可能なプロパティとその定義域および値域を獲得できる.

6 の獲得した既存 OWL オントロジーのランキングについては, 5. 章で述べる.

*3 [1] で指定可能な検索キーワードについては, 以下の URL が参考になる.
http://swoogle.umbc.edu/index.php?option=com_swoogle_manual&manual=search_swd

5. 既存 OWL オントロジーのランキング

DODDLE-OWL は、現状では Swoogle の OntoRank 及び TermRank, オントロジー中の入力概念を含む割合, オントロジー中の入力概念に関する非階層関係の数の 4 つをランキングのための指標としている。OntoRank 及び TermRank は、それぞれ、オントロジーと SWT が Swoogle が収集した SWD 集合の中で、どのように参照されているかを解析し、ランキングした値である。OntoRank 及び TermRank を求める式は、[Ding 05] に定義されている。DODDLE-OWL では、入力概念をより多く含むオントロジーは対象領域に関連するオントロジーとみなす。また、入力概念に関連する非階層関係をより多く定義しているオントロジーも対象領域に関連するオントロジーとみなす。入力概念を含む割合が同程度のオントロジーについては、OntoRank を参考にする事で、ユーザはより多くの SWD で参照されているオントロジーを再利用することが可能となる。入力概念の候補が複数ある場合には、TermRank を参考にする事で、より多くの SWO で参照されている入力概念を再利用することが可能となる。

6. 既存 OWL オントロジーからの要素抽出

OWL オントロジーを参照オントロジーとして領域オントロジー構築支援で利用するためには、オントロジー構築支援に利用可能な要素を OWL オントロジーから抽出する必要がある。DODDLE-OWL では、領域オントロジーにおける階層関係および非階層関係の定義を支援している。概念関係および非階層関係を定義するために必要な OWL オントロジーの要素は、概念(クラス及びプロパティ)、概念の見出し、概念の説明、上位・下位関係、非階層関係である。概念の抽出は、領域オントロジー構築において必須である。概念の見出しは、入力単語と概念を対応づける(入力概念選択)ために必要である。概念の説明は、入力単語に対応する概念が複数ある場合に、入力概念をユーザが選択する際の参考となる。上位・下位関係はクラス及びプロパティ階層を構築する際に必要である。非階層関係を定義するために、プロパティの定義域および値域などを抽出する必要がある。

RDFS, DAML, OWL などのオントロジー記述言語は、上記のオントロジーの要素を定義するために基本的なクラスやプロパティを提供している。Swoogle では、現状では RDFS, DAML, OWL が提供している基本的なクラス及びプロパティに基づいて、オントロジーの検索を行うことができるようになっている。例えば、Swoogle ではクラスを、次の (X, Y, Z) という Statement を満たす X と定義している。

- X: 匿名以外のリソース
- Y: rdf:type
- Z: rdfs:Class, owl:Class, owl:Restriction, owl:DataRange, daml:Class, daml:Datatype, daml:Restriction のいずれか

Swoogle が扱う範囲のクラスやプロパティのみを OWL オントロジーから抽出する場合には、Swoogle の定義に従ってオントロジーの要素を抽出すればよい。しかし、それ以外の形式で定義された汎用オントロジーやシソーラスが存在する。[小出 06] では WordNet や EDR の OWL 化について検討し

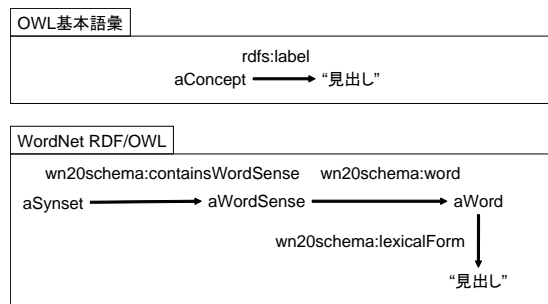


図 2: OWL 基本語彙および WordNet RDF/OWL における概念の見出し

ており、Swoogle で定義しているクラスやプロパティとは構造が異なっている。[中山 06] では、Wikipedia に対して Web マイニングを行う手法である Wikipedia マイニングを提案し、シソーラス辞書 *4 を構築している。構築された Wikipedia シソーラスを OWL 形式で表現するための語彙として、SKOS (Simple Knowledge Organisation System) *5 を用いている。SKOS では、概念を表すクラスとして skos:Concept, 上位概念を表すプロパティとして skos:broader などが定義されており、OWL 基本語彙とは異なっている。図 2 に OWL 基本語彙及び WordNet RDF/OWL における概念の見出しを示す。表 2 に OWL 基本語彙, SKOS, WordNet RDF/OWL, EDR RDF/OWL におけるオントロジーの要素を特定するクラス及びプロパティを示す。

DODDLE-OWL では、多様な形式のオントロジーからオントロジーの要素を抽出するために、クラス抽出テンプレート、プロパティ抽出テンプレート、見出しと説明抽出テンプレート、上位・下位関係抽出テンプレート、非階層関係抽出テンプレートの 5 種類のテンプレートを RDF クエリー言語 SPARQL *6 を用いて記述し、OWL オントロジーと対応づけている。図 3 から図 7 に OWL 語彙におけるオントロジーの要素を抽出するための SPARQL で記述したテンプレートを示す。

図 5 の見出しと説明抽出テンプレートを直接 SPARQL のクエリーとした場合、OWL オントロジー中のすべての rdfs:label と rdfs:comment プロパティの値を抽出してしまう。DODDLE-OWL は ?concept の部分を取得したい概念(クラスまたはプロパティ)の URI に置換することで、特定の概念の見出し及び説明のみを抽出できるようにしている。他のテンプレートも同様にテンプレートを直接 SPARQL のクエリーとして用いるのではなく、変数部分を DODDLE-OWL プログラム中で適切な URI に置換したものを最終的な SPARQL のクエリーとしている。?concept, ?subConcept, ?class, ?property, ?label, ?description, ?domain, ?range 変数を用いてトリプルのパターンを各オントロジーの要素を抽出するテンプレートに記述し、テンプレートを OWL オントロジーに対応づけることで、様々なクラス、プロパティ、構造により表現されたオントロジーの要素を抽出することが可能となる。

*4 <http://wikipedia-lab.org/>

*5 <http://www.w3.org/TR/swbp-skos-core-guide/>

*6 <http://www.w3.org/TR/rdf-sparql-query/>

表 2: OWL 基本語彙, SKOS, WordNet RDF/OWL, EDR RDF/OWL におけるオントロジーの要素を特定するクラス及びプロパティ

| オントロジーの要素 | オントロジーの要素を特定するクラス及びプロパティ |
|-----------|--|
| 概念 | rdfs:Class, owl:Class, rdf:Property, owl:ObjectProperty, owl:DatatypeProperty skos:Concept wn20schema:WordSense, wn20schema:NounWordSense など edrschema:概念 |
| 見出し | rdfs:label skos:prefLabel, skos:altLabel, skos:hiddenLabel wn20schema:lexicalForm edrschema:日本語概念見出し, edrschema:英語概念見出し |
| 説明 | rdfs:comment skos:definition wn20schema:gloss edrschema:日本語概念説明, edrschema:英語概念説明 |
| 上位・下位関係 | rdfs:subClassOf, rdfs:subPropertyOf skos:broader, skos:narrower wn20schema:hypernymOf, wn20schema:hyponymOf edrschema:上位概念, edrschema:下位概念 |
| 非階層関係 | rdfs:domain, rdfs:range skos:related wn20schema:antonymOf, wn20schema:partMeronymOf など edrschema:agent, edrschema:object |

rdf, rdfs, owl, skos, wn20schema, edrschema はそれぞれ以下の名前空間の接頭辞を表している . rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#, rdfs=http://www.w3.org/2000/01/rdf-schema#, owl=http://www.w3.org/2002/07/owl#, skos=http://www.w3.org/2004/02/skos/core#, wn20schema=http://www.w3.org/2006/03/wn/wn20/schema/, edrschema=http://www.kasm.nii.ac.jp/ontologies/edr/Schema#

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?class
WHERE {
  {?class rdf:type owl:Class} UNION {?class rdf:type rdfs:Class}
}
```

図 3: OWL 語彙におけるクラス抽出テンプレート

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?property
WHERE {
  {?property rdf:type owl:ObjectProperty} UNION
  {?property rdf:type owl:DatatypeProperty} UNION
  {?property rdf:type rdf:Property}
}
```

図 4: OWL 語彙におけるプロパティ抽出テンプレート

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?label ?description
WHERE {
  {?concept rdfs:label ?label} UNION
  {?concept rdfs:comment ?description}
}
```

図 5: OWL 語彙における見出しと説明抽出テンプレート

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?subConcept
WHERE {
  {?subConcept rdfs:subClassOf ?concept} UNION
  {?subConcept rdfs:subPropertyOf ?concept}
}
```

図 6: OWL 語彙における上位・下位関係抽出テンプレート

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?property ?domain ?range
WHERE {
  {?property rdfs:domain ?domain} UNION
  {?property rdfs:range ?range}
}
```

図 7: OWL 語彙における非階層関係抽出テンプレート

表 3: CISG 第二部における主要 46 概念

| | | | | | | | |
|------|----------------------|-----|-------------|------|-------------------|-------|--------------|
| 承諾 | acceptance | 遅延 | delay | 修正 | modification | 量 | quantity |
| 行為 | act | 配達 | delivery | 申込 | offer | 拒絶 | rejection |
| 付加 | addition | 不一致 | discrepancy | 申込者 | offerer | 回答 | reply |
| 住所 | address | 発送 | dispatch | 被申込者 | offeree | 居住地 | residence |
| 同意 | assent | 効果 | effect | 当事者 | party | 取消 | revocation |
| 状況 | circumstance | 書簡 | envelope | 支払 | payment | 沈黙 | silence |
| 通信 | communication system | 物品 | goods | 人 | person | 陳述 | speech act |
| | | 休日 | holiday | | | 電話 | telephone |
| 行為 | conduct | 表示 | indication | 営業所 | place of business | テレックス | telex |
| 契約 | contract | 意志 | intention | 価格 | price | 時間 | time |
| 反対申込 | counteroffer | 誘引 | invitation | 申入 | proposal | 伝達 | transmission |
| 日 | day | 手紙 | letter | 質 | quality | 撤回 | withdrawal |

7. ケーススタディ

表 4: [3] により獲得したクラス数, プロパティ数, 入力単語のカバー率

| | |
|-----------|---------------|
| クラス数 | 141 |
| プロパティ数 | 1 |
| 入力単語のカバー率 | 0.717 (33/46) |

表 5: [17] 及び [19] により獲得したプロパティ数と非階層関係数

| | |
|--------|-----|
| プロパティ数 | 495 |
| 非階層関係数 | 47 |

表 6: [16] 及び [18] により獲得した非階層関係数

| | |
|--------|--------|
| 非階層関係数 | 111209 |
|--------|--------|

4. 5. 6. 章で述べた手法を用いて, Swoogle を用いた既存 OWL オントロジーの獲得, ランキング, 概念定義支援を行った. 本ケーススタディでは, 入力文書として法律分野の専門文書である国際売買法 [Sono 93] 第二部の英文テキスト (以下, CISG) を用いた. CISG 第二部に関する概念選択, 階層構築, 概念定義は, [M.Kurematsu 04] において, 法律の専門家によってなされているため, 本ケーススタディではそれらを正解としている. CISG 第二部における主要 46 概念を表 3 に示す. 本ケーススタディは以下に示す手順で行った.

手順 1 では, 表 3 に示した CISG 第二部における主要 46 概念を入力として, 表 1 の [3] を用いて, 入力単語をローカル名または `rdfs:label` プロパティの値とするクラス及びプロパティを獲得した. 本ケーススタディでは, [3] を用いて TermRank

表 7: [1], [4], [13] を用いて獲得したオントロジー数およびオントロジーに定義されている入力概念に関連するクラス数, プロパティ数, 非階層関係数

| | |
|---------|-----|
| オントロジー数 | 105 |
| クラス数 | 141 |
| プロパティ数 | 490 |
| 非階層関係数 | 0 |

表 8: 定義域および値域が入力概念または入力概念の上位概念であるプロパティ数および非階層関係数

| | |
|--------|------|
| プロパティ数 | 490 |
| 非階層関係数 | 1149 |

表 9: Swoogle を用いて獲得した OWL オントロジーより得られた非階層関係

| 定義域 | 値域 | 厳密照合数 | 拡張照合数 |
|----------|----------|-------|-------|
| person | person | 202 | 891 |
| person | address | 2 | 1 |
| reply | person | 2 | 0 |
| letter | person | 1 | 3 |
| person | letter | 1 | 3 |
| address | person | 1 | 1 |
| day | person | 0 | 18 |
| quantity | person | 0 | 18 |
| person | day | 0 | 14 |
| person | quantity | 0 | 14 |
| day | day | 0 | 5 |
| day | quantity | 0 | 5 |
| quantity | day | 0 | 5 |
| quantity | quantity | 0 | 5 |
| proposal | proposal | 0 | 4 |

によりランキングされた、上位 5 つまでのクラスおよびプロパティを獲得した。表 4 に [3] により獲得したクラス数、プロパティ数、入力単語のカバー率を示す。

手順 2 では、表 1 の [17] 及び [19] を用いて、手順 1 で獲得したクラスを定義域または値域とするプロパティを獲得した。本ケーススタディでは、[17] 及び [19] の検索結果のうち最初の 100 個までのプロパティを獲得した。表 5 に [17] 及び [19] により獲得したプロパティ数と、手順 2 における非階層関係数を示す。非階層関係数は、プロパティにおける定義域の数と値域の数の積であらわされる。

手順 3 では、表 1 の [16] 及び [18] を用いて、手順 2 で獲得したプロパティの定義域および値域を獲得した。本ケーススタディでは、[16] 及び [18] の検索結果のうち最初の 100 個までの定義域および値域を用いた。表 6 に [16] 及び [18] により獲得した非階層関係数を示す。

手順 4 では、表 1 の [1], [4], [13] を用いて、手順 1 から 3 で獲得したクラス及びプロパティを定義しているオントロジーを獲得した。はじめに、[1] を用いて OntoRank によりランキングされた、各入力単語に関連する上位 10 個のオントロジーを獲得した。手順 1 から 3 で獲得したクラス及びプロパティが [1] を用いて獲得したオントロジーに含まれていない場合は、[4] 及び [13] を用いて個別にオントロジーを獲得した。表 7 に [1], [4], [13] を用いて獲得したオントロジー数およびオントロジーに定義されている入力概念に関連するクラス数、プロパティ数、非階層関係数を示す。

手順 5 では、手順 1 及び 2 で獲得したプロパティの中で、定義域および値域が入力概念または入力概念の上位概念でないものを削除した。表 8 に定義域および値域が入力概念または入力概念の上位概念であるプロパティ数および非階層関係数を示す。表 9 に Swoogle を用いて獲得した OWL オントロジーより得られた非階層関係を示す。表 9 では、定義域および値域を対応する入力単語に置換した場合の非階層関係を示している。表 9 の厳密照合数は定義域と値域が共に、入力概念の場合の非階層関係数をあらわす。表 9 の拡張照合数は定義域または値域のどちらかが、入力概念の上位概念の場合の非階層関係数をあらわす。

手順 6 では、Swoogle の OntoRank、オントロジー中の入力概念を含む割合、オントロジー中の入力概念に関する非階層関係の数をランキングのための指標とした。表 10 に、Swoogle を用いて獲得した、オントロジー中の入力概念を含む割合でソートした上位 20 個の OWL オントロジーを示す。本ケーススタディでは、入力概念に関する非階層関係が得られなかったため、表 10 には載せていない。

7.1 考察

表 4 より、約 7 割の入力単語については、入力単語を表記として持つ概念を得ることができた。しかし、表 10 より、入力概念を含む割合が最も大きいオントロジーで 3 割程度であった。また、入力概念を多く含むオントロジーの大部分は、WordNet、FrameNet^{*7}、Cyc^{*8}などの汎用オントロジーであり、領域オントロジーではなかった。獲得されたオントロジーの大部分が汎用オントロジーであった原因として、手順 1 で TermRank の高いクラス及びプロパティを獲得したことが考えられる。TermRank の高いクラスは、様々なオントロジーや

RDF から参照されているため、汎用的であると考えられる。手順 1 で獲得した汎用的なクラスやプロパティを基にしたため、最終的に汎用的なオントロジーを多く獲得してしまったと考えられる。

表 9 に示した非階層関係は、[M.Kurematsu 04] で法律の専門家が定義した関係を含んでいなかった。これも、獲得されたオントロジーの中に、汎用オントロジーが多かったため、一般的な関係が多く抽出されてしまったことが原因だと考えられる。

表 7 より、獲得したオントロジーの中に入力概念に関連する非階層関係の定義を得ることができなかった。これは、表 1 の [16] 及び [18] により得られるプロパティの定義域および値域は、複数のオントロジーから獲得しているため、定義域と値域が共に 1 つのオントロジーの中に定義されているとは限らないことが考えられる。また、本ケーススタディでは、表 1 の [1] を用いて、上位 10 個のオントロジーのみを獲得したため、すべてのオントロジーを獲得した場合には、1 つのオントロジー中で入力概念に関連する非階層関係の定義が得られると思われる。

Swoogle Web サービスにより獲得できるオントロジーには類似したものが多く含まれていることが本ケーススタディを通してわかった。表 10 より、内容はほぼ同様で URL のみが異なるオントロジーが多く含まれていることがわかる。Swoogle を用いて獲得したオントロジーの中から、ユーザが領域に関連するオントロジーを選択する際には、同じ内容のオントロジーについては一つにまとめて提示できた方がよいと思われる。本ケーススタディでは、獲得した OWL オントロジーのメタデータ（ヘッダ）については考慮していなかったが、バージョン情報などがあれば、バージョン違いのオントロジーを同定することができると考えられる。また、OWL メタデータがない場合には、URL の類似度やオントロジーに定義されているクラスやプロパティの類似度などから、内容の近いオントロジーを同定することができると考えられる。

8. おわりに

本稿では、オントロジー検索エンジン Swoogle を用いた領域オントロジー構築支援環境 DODDLE-OWL の拡張について述べた。Swoogle を用いて獲得した OWL オントロジー及び専門文書を参照リソースとした、領域オントロジー構築支援を行うために、Swoogle を用いた既存 OWL オントロジーの獲得、既存 OWL オントロジーのランキング、既存 OWL オントロジーからの要素抽出の 3 つの拡張を行った。ケーススタディとして、法律分野の専門文書である国際売買法 [Sono 93] 第二部の英文テキストを入力として、提案手法により既存 OWL オントロジーの獲得および概念定義支援を行った。領域にとって重要な単語（入力単語）を表記として持つクラスを獲得する際に、Swoogle [Ding 05] の TermRank の高いクラスを獲得したために、最終的に汎用オントロジーが数多く獲得された。

今後の課題として、入力単語を表記として持つクラスを獲得する際に、Swoogle の TermRank とは異なる指標を用いて、領域に関連するクラスを獲得する方法を考える必要がある。また、複数の参照オントロジーから関連する概念を抽出、合成して領域オントロジーを構築する場合には、上位概念階層の構造の違いにより単純に合成することは困難であると考えられる。その際には、オントロジーアライメントにより類似概念の同定

*7 <http://framenet.icsi.berkeley.edu/>

*8 <http://www.cyc.com/>

表 10: Swoogle を用いて獲得した OWL オントロジー（オントロジー中の入力概念を含む割合でソートした上位 20 個）

| 順位 | OntoRank | 入力概念を含む割合 (入力概念数) | オントロジーの URL |
|----|----------|-------------------|---|
| 1 | 0.881 | 0.326 (15) | http://www.w3.org/2001/sw/BestPractices/WNET/wNounsyn_v7.owl |
| 2 | 0.881 | 0.326 (15) | http://www.loa-cnr.it/ontologies/OWN/OWN.owl |
| 3 | 642.815 | 0.261 (12) | http://morpheus.cs.umbc.edu/aksl/ontosem.owl |
| 4 | 0.860 | 0.261 (12) | http://www.cs.umbc.edu/aksl/ontosem.owl |
| 5 | 0.735 | 0.239 (11) | http://rhizomik.net/ontologies/2005/07/FrameNet_1.1_inferred.owl |
| 6 | 0.726 | 0.239 (11) | http://rhizomik.upf.edu/ontologies/2005/07/FrameNet_1.1.owl |
| 7 | 0.726 | 0.239 (11) | http://rhizomik.net/ontologies/2005/07/FrameNet_1.1.owl |
| 8 | 0.875 | 0.217 (10) | http://athena.ics.forth.gr:9090/RDF/VRP/Examples/DCD100.rdf |
| 9 | 0.875 | 0.217 (10) | http://139.91.183.30:9090/RDF/VRP/Examples/DCD100.rdf |
| 10 | 1.701 | 0.196 (9) | http://www.cyc.com/2004/06/04/cyc |
| 11 | 5.725 | 0.174 (8) | http://semweb.mcdonaldbradley.com/OWL/Cyc/FreeToGov/060704/FreeToGovCyc.owl |
| 12 | 0.924 | 0.174 (8) | http://semweb.mcdonaldbradley.com/dev/OWL/Cyc/FreeToGov/060704/FreeToGovCyc.owl |
| 13 | 0.924 | 0.174 (8) | http://semweb.mcdonaldbradley.com/Private/OWL/Cyc/FreeToGov/060704/FreeToGovCyc.owl |
| 14 | 1.280 | 0.152 (7) | http://ontologies.isx.com/onts/saturn/2004/10/core.owl |
| 15 | 0.875 | 0.152 (7) | http://www.cs.utexas.edu/mfkb/RKF/tree/CLib-core-office.owl |
| 16 | 0.875 | 0.152 (7) | http://www.cs.utexas.edu/users/mfkb/RKF/tree/CLib-core-office.owl |
| 17 | 0.837 | 0.152 (7) | http://dmag.upf.es/ontologies/2003/12/ipronto.owl |
| 18 | 0.837 | 0.152 (7) | http://dmag.upf.edu/ontologies/2003/12/ipronto.owl |
| 19 | 1.363 | 0.130 (6) | http://co4.inrialpes.fr/align/Contest/223/onto.rdf |
| 20 | 1.792 | 0.109 (5) | http://counterterror.mindswap.org/2005/terrorism.owl |

が必要になると考えられるが、オントロジーアライメントを用いた階層構築支援は今後の課題である。

参考文献

[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/>

[M.Kurematsu 04] M.Kurematsu, , T.Iwade, , N.Nakaya, , and T.Yamaguchi, : DODDLE II A Domain Ontology Development Environment Using a MRD and Text Corpus, *IEICE(E)*, pp. 908–916 (2004)

[Morita 06] Morita, T., Fukuta, N., Izumi, N., and Yamaguchi, T.: DODDLE-OWL: A Domain Ontology Construction Tool with OWL, *First Asian Semantic Web Conference, LNCS4185*, pp. 537–551 (2006)

[Sono 93] Sono, K. and Yamate, M.: *United Nations Convention on Contracts for the International Sale of Goods*, Seirin Shoin (1993)

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

[Yokoi 95] Yokoi, T.: The EDR electronic dictionary, *Commun. ACM*, Vol. 38, No. 11, pp. 42–44 (1995)

[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)

[小出 06] 小出 誠二, 森田 武史, 山口 高平, ムリアディヘンドリー, 武田 英明: WordNet と EDR の OWL 表現, *人工知能学会セマンティック Web とオントロジー研究会 SIG-SWO-A601-03* (2006)

[森田 06] 森田 武史, 山口 高平: 日本語概念を対象にした領域オントロジー構築支援環境 DODDLE の機能拡張, *電子*

情報通信学会, 信学技報 (知能ソフトウェア工学研究会) KBSE2006-1 (2006)

[中山 06] 中山 浩太郎, 原 隆浩, 西尾 章治郎: Wikipedia マイニングによるシソーラス辞書構築手法, *情報処理学会論文誌*, Vol. 47, No. 10, pp. 2917–2928 (2006), <http://wikipedia-lab.org/>