

# オントロジー検索エンジンと 領域オントロジー構築支援環境 DODDLE-OWL の統合

## Integrating DODDLE-OWL and an Ontology Search Engine

森田 武史\*<sup>1</sup>      和泉 憲明\*<sup>2</sup>      山口 高平\*<sup>1</sup>  
Takeshi Morita      Noriaki Izumi      Takahira Yamaguchi

\*<sup>1</sup>慶應義塾大学  
Keio University

\*<sup>2</sup>独立行政法人 産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology

In order to reduce the costs of building up domain ontologies manually, in this paper, we integrate a domain ontology development environment: DODDLE-OWL and an ontology search engine: Swoogle. We propose a domain ontology construction method reusing texts and existing ontologies extracted by Swoogle. In this evaluation, we applied the method to a particular field of law and evaluated the acquired ontologies.

### 1. はじめに

セマンティック Web を実現するためには、機械処理可能なメタデータを定義するためのオントロジーが必要である。しかし、特定の領域概念を扱う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、開発コストが高いという問題がある。

先行研究において、汎用オントロジーとテキストを参照リソースとした領域オントロジー構築支援環境 DODDLE-OWL [Morita 06] を開発してきた。DODDLE-OWL では、Web 上に存在する OWL オントロジー（既存オントロジー）の再利用ができなかった。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り、より容易に領域オントロジーが構築できると考えられる。また、既存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、キーワードを含むオントロジー、クラス、プロパティ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参照しながら、目的に合った領域オントロジーを手動で構築する必要があり、その作業にはコストがかかる。

本研究では既存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オントロジー検索エンジンと DODDLE-OWL の統合を行う。また、オントロジー検索エンジンを用いて獲得した既存オントロジー及びテキストを参照リソースとした、領域オントロジー構築支援手法及びその実証実験について述べる。

### 2. オントロジー検索エンジン

既存のオントロジー検索エンジンとして、Swoogle [Ding 05] がある。Swoogle には、2007 年 4 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を検索することができる。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) も提供されている。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、Web サービスを提供している Swoogle を DODDLE-OWL では利用する。

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既存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

#### 3.1 領域オントロジー構築支援環境の設計

図 1 に Swoogle を用いた 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 (階層関係および非階層関係) である。はじめに、形態素解析器を用いて領域専門文書から名詞及び動詞を抽出する。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) をユーザは選択する。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、獲得された既存オントロジーの中から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) をユーザは選択する。領域によっては既存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、既存オントロジーが獲得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

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

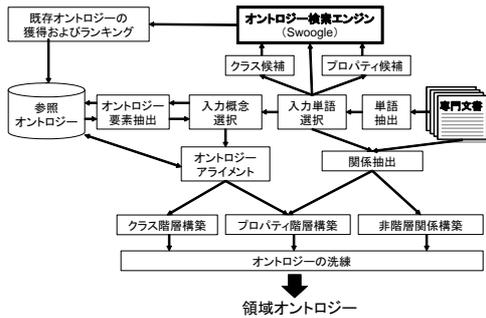


図 1: Swoogle を用いた DODDLE-OWL のシステムフロー

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。参照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、1. Swoogle を用いた既存オントロジー獲得、2. 既存オントロジーのランキング、3. 既存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の主に 4 つの拡張を行う必要がある。以下では上記 1 及び 2 について述べる。

### 3.2 Swoogle を用いた既存オントロジー獲得

以下の手順により、Swoogle を用いて DODDLE-OWL は既存オントロジーの獲得を行う。実装の詳細は、4. 章で述べる。

1. 入力単語を URI のローカル名または `rdfs:label` プロパティの値として持つクラス及びプロパティを獲得
2. 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
3. 手順 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
4. 手順 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
5. 手順 1 と 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

### 3.3 既存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、[Ding 05] で提案されているオントロジーとクラス及びプロパティをランキングするための指標である。本研究では、入力概念をより多く含むオントロジーを対象領域に関連するオントロジーと仮定する。また、入力概念に関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仮定する。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セマンティック Web 文書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。語の多義性により、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、ユーザはより多くのオントロジー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。

## 4. Swoogle を用いた既存オントロジー獲得の実装

本章では、3.2 節で述べた、Swoogle を用いた既存オントロジー獲得の実装について述べる。Swoogle は 19 種類の REST 形式の Web サービスを提供している。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。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 の SWT (Semantic Web Term) はクラスまたはプロパティを表す。SWD (Semantic Web Document) は RDF/XML, N-Triple, N3 形式で記述された RDF 文書を表す。SWO (Semantic Web Ontology) 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。

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などを与える。`key` は出力する検索結果数に制限をかけるために用いる。

以下では、3.2 節で述べた手順の詳細について述べる。

手順 1 では表 1 の [3] を用いて、入力単語を URI のローカル名または `rdfs:label` プロパティの値として持つクラス及びプロパティを獲得する。例えば、`searchString` パラメータに (`localname:person OR label:person`) (`type:owl.class OR type:rdfs.class`) を指定することにより、`person` を URI のローカル名または `rdfs:label` プロパティの値として持つクラスを獲得できる。また、[3] では、獲得した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。

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する。

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する。

手順 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、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できる。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、URI のローカル名、`rdfs:label` プロパティの値などに、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、手順 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、

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

[1] で獲得した上位 25 個のオントロジーの中に手順 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.

手順 5 では, 手順 4 で獲得したオントロジーを参照して, 定義域および値域が共に入力概念または入力概念の上位概念でないプロパティを削除する. 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.

## 5. 実証実験

### 5.1 実験の概要

表 2: 獲得したオントロジー数, 入力単語の網羅率, 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

獲得したオントロジー数	120
入力単語の網羅率	0.717 (33/46)
クラス数	331
プロパティ数	558

表 3: 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非階層関係数, 厳密照合及び拡張照合した非階層関係の種類の数

非階層関係数	1744
厳密照合した非階層関係の種類の数	7
拡張照合した非階層関係の種類の数	182

3. 章および 4. 章で述べた手法を用いて, Swoogle を用いた既存オントロジーの獲得, ランキング, 概念定義支援を行った. 本実験では, 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 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, 本実験ではそれらを正解としている. 本実験は以下に示す手順で行った.

表 4: 厳密照合した非階層関係の種類

定義域	値域	厳密照合数
person	person	94
person	address	3
letter	person	2
reply	person	2
address	person	1
telephone	effect	1
person	letter	1

表 5: 専門家の非階層関係定義と一致した非階層関係の種類

定義域	値域	厳密照合数	拡張照合数
quantity	goods	0	41
contract	offer	0	40
holiday	holiday	0	40
offer	goods	0	40
letter	envelope	0	3
reply	offer	0	1
time	contract	0	1
time	offer	0	1

手順 1 では, CISG 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 `rdfls:label` プロパティの値とする, TermRank によりランキングされた上位 5 つまでのクラス及びプロパティを獲得した.

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. 本実験では, 検索結果の上位 100 個までのプロパティを獲得した.

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. 本実験では, 検索結果の上位 100 個までの定義域および値域を獲得した.

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 2 に獲得したオントロジー数, 入力単語の網羅率, 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, プロパティ数を示す.

手順 5 では, 手順 4 で獲得したオントロジーを参照して, 定義域および値域が共に入力概念または入力概念の上位概念でないプロパティを削除した. 表 3 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非階層関係数, 厳密照合及び拡張照合した非階層関係の種類の数を示す. ここでは, 定義域と値域が共に入力概念である非階層関係を厳密照合した非階

表 6: Swoogle を用いて獲得した既存オントロジー (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 10 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	0.348 (16)	<a href="http://www.loa-cnr.it/ontologies/OWN/OWN.owl">http://www.loa-cnr.it/ontologies/OWN/OWN.owl</a>
2	0.881	0.326 (15)	<a href="http://www.w3.org/2001/sw/BestPractices/WNET/wnNounsyn_v7.owl">http://www.w3.org/2001/sw/BestPractices/WNET/wnNounsyn_v7.owl</a>
3	642.815	0.261 (12)	<a href="http://morpheus.cs.umbc.edu/aks1/ontosem.owl">http://morpheus.cs.umbc.edu/aks1/ontosem.owl</a>
4	0.860	0.261 (12)	<a href="http://www.cs.umbc.edu/aks1/ontosem.owl">http://www.cs.umbc.edu/aks1/ontosem.owl</a>
5	0.735	0.239 (11)	<a href="http://rhizomik.net/ontologies/2005/07/FrameNet_1.1_inferred.owl">http://rhizomik.net/ontologies/2005/07/FrameNet_1.1_inferred.owl</a>
6	0.726	0.239 (11)	<a href="http://rhizomik.upf.edu/ontologies/2005/07/FrameNet_1.1.owl">http://rhizomik.upf.edu/ontologies/2005/07/FrameNet_1.1.owl</a>
7	0.726	0.239 (11)	<a href="http://rhizomik.net/ontologies/2005/07/FrameNet_1.1.owl">http://rhizomik.net/ontologies/2005/07/FrameNet_1.1.owl</a>
8	0.875	0.217 (10)	<a href="http://athena.ics.forth.gr:9090/RDF/VRP/Examples/DCD100.rdf">http://athena.ics.forth.gr:9090/RDF/VRP/Examples/DCD100.rdf</a>
9	0.875	0.217 (10)	<a href="http://139.91.183.30:9090/RDF/VRP/Examples/DCD100.rdf">http://139.91.183.30:9090/RDF/VRP/Examples/DCD100.rdf</a>
10	1.701	0.196 (9)	<a href="http://www.cyc.com/2004/06/04/cyc">http://www.cyc.com/2004/06/04/cyc</a>

層関係とする。また、定義域または値域のどちらかが、入力概念の上位概念である非階層関係を拡張照合した非階層関係とする。表 3 の厳密照合した非階層関係の種類の数とは、定義域および値域に対応する入力単語に置換した場合の入力単語対の数をあらわす。拡張照合した非階層関係の種類の数も同様である。表 4 に厳密照合した非階層関係の種類を示す。表 5 に専門家の非階層関係定義と一致した非階層関係の種類を示す。

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 6 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 10 個の既存オントロジーを示す。

## 5.2 実験結果の検討

表 2 より、CISG の約 7 割の入力単語については、獲得したオントロジーで網羅することができる。しかし、表 6 より、入力概念を含む割合が最も多い既存オントロジーで 3 割程度であり、入力概念をすべて含む領域オントロジーを構築するためには、複数の既存オントロジーを参照する必要があることがわかる。また、入力概念を多く含むオントロジーの大部分は、WordNet \*1 や Cyc \*2 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。手順 1 で獲得した TermRank の高い、汎用的なクラス及び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 6 より、内容は等しいが URL のみが異なる既存オントロジーが多く獲得された。Swoogle を用いて獲得した既存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、内容の等しい既存オントロジーを一つにまとめて提示できるとよい。本実験では、owl:imports プロパティ以外の既存オントロジーにおける OWL メタデータを利用していない。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、内容の等しいオントロジーを同定できると考えられる。

[M.Kurematsu 04] において専門家は 44 の非階層関係を定義していたが、本実験ではそのうち、表 5 に示した 8 つの非階層関係を獲得できた。表 5 に示した非階層関係はすべて拡張照合であった。表 4 に示した厳密照合した非階層関係の種類は、person との関係が多かった。汎用オントロジーを多く獲得したために、一般的な関係が多く抽出されたことが原因だと考えられる。

表 3 より、本実験では約 1700 という膨大な数の非階層関係を獲得した。これらの中からユーザが必要な非階層関係を選択

することは困難であるため、獲得した非階層関係のランキングや類似した非階層関係をまとめて提示するための機能が必要になると考えられる。

## 6. おわりに

本稿では、オントロジー検索エンジン Swoogle と領域オントロジー構築支援環境 DODDLE-OWL の統合について述べた。Swoogle を用いて獲得した既存オントロジー及びテキスト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 を用いた既存オントロジーの獲得、既存オントロジーのランキング、既存オントロジーからの要素抽出の 3 つの拡張を DODDLE-OWL に行った。実証実験として、法律分野の専門文書を入力として、提案手法により既存オントロジーの獲得および概念定義支援を行った。Swoogle の TermRank の高い汎用的なクラスを基にしたため、獲得された既存オントロジーの大部分は汎用オントロジーであった。また、得られた概念定義も一般的な関係が多かった。

今後の課題として、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、オントロジーアライメントを用いた、複数の既存オントロジーからの階層構築支援も今後の課題である。

## 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, in *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, in *Proceedings of the First Asian Semantic Web Conference, LNCS 4185*, 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)

\*1 <http://wordnet.princeton.edu/>

\*2 <http://www.cyc.com/>