1G3-01

セマンティックWebのためのオントロジー構築支援環境 DODDLE-Rの実装

DODDLE-R: A Domain Ontology Development Environment for the Semantic Web

森田武史[*1] 繁田佳宏[*1] 杉浦直樹[*1] 福田直樹[*2] 和泉憲明[*3] 山口高平[*4]
Takeshi Morita Yoshihiro Shigeta Naoki Sugiura Naoki Fukuta Noriaki Izumi Takahira Yamaguchi

[*1] 静岡大学大学院情報学研究科 Graduate School of Informatics, Shizuoka University
[*2] 静岡大学情報学部 Faculty of Informatics, Shizuoka University
[*3] 産業技術総合研究所 サイバーアシスト研究センター Cyber Assist Research Center, National Institute of Advanced Industrial Science and Technology
[*4] 慶應義塾大学理工学部 Faculty of Science and Technology, Keio University

In this paper, we propose an ontology development environment for the Semantic Web. The advantage of our environment is focusing the quality improvement phase of ontology construction. The environment supports the user by two steps. First, an initial ontology is generated semi-automatically. Then the environment supports the user to refine the ontology by providing candidates of relationships to be added and modified using syntactic analysis and domain specific taxonomy. Furthermore, bugs of ontologies are made easy to discover by integrating the environment and a meta-model management tool. Through interactive support for improving the quality of initial ontology, OWL-Lite level ontology, which consists of taxonomic relationships (class - sub class relationship) and non-taxonomic relationships (defined as property), is constructed effectively. However, There are some problems to explort an ontology in OWL syntax. We discuss about the problems and possible solutions.

1. はじめに

セマンティックWebにおける意味レベルでの情報検索など，計算機による意味解釈が必要な際には，オントロジーが不可欠である．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が膨大であること，領域概念の専門性の高さなど構築コストの高さがボトルネックとなっている．本稿では，セマンティックWebにおける領域オントロジーの開発コストを軽減することを目的とした，セマンティックWebのための領域オントロジー構築支援環境を提案する．

2. 領域オントロジー構築支援環境

領域オントロジー構築支援環境DODDLE-R(Domain Ontology rapiD Development Environment - RDF [RDF] Extension)のシステムフローを図1に示す．システムへの入力として，オントロジーを構成する語彙集合，参照として電子化辞書とテキストコーパスを用いて，出力として領域オントロジーを獲得する．DODDLE-Rは先行研究のDODDLE-II [Kurematsu 04]にオントロジー洗練モジュールを追加し，メタモデル管理ツール(MR^3: Meta-Model Management based on RDFs Revision Reflection) [Morita 03]との相互運用によりオントロジー構築過程を視覚化し，最終的に領域オントロジーをOWL(Web OntologyLanguage) [OWL]形式でエクスポートする．以下，各モジュールについて述べる．

2.1 オントロジー構築モジュール

オントロジー構築モジュールでは，入力語彙に基づいてWordNet [G.A.Miller 95]を参照し，入力概念を得る．入力概念を含む部分木をWordNetから抽出し，概念階層の雛型である**初期モデル**を構築する．テキストコーパスから共起性に基づく統計処理であるWordSpaceと相関ルール[R.Agrawal 94]により，概念定義に関わる可能性のある**概念対**を抽出し，それらの関係値を得る．

2.2 オントロジー洗練モジュール

オントロジー洗練モジュールでは，オントロジー構築モジュールで生成された初期モデルに領域依存部分の修正を加えるための支援を行う．オントロジー構築モジュールで得られた概念対集合に，共起性に基づく統計処理による重み付けを行う．先行研究に加えて構文情報と概念階層を利用した重みづけも行い，評価値にしたがい概念仕様テンプレートを構築し，ユーザに提示する．ユーザは概念仕様テンプレートを基に概念定義を行う．最終的に概念階層と概念定義を合成し，領域オントロジーを得る．以下，構文情報および概念階層を利用した重み付けについて説明する．

構文情報の利用

共起性に構文情報を加味した概念対の重みづけを行う．テキストコーパスの文中から抽出した概念対には，文構造を見た場合，主部・述部の関係にあるものが存在する．このような概念対に対して，一般的な概念対よりも関係が深い可能性が高いものと仮定し，関係値を算出する．重みづけを行うために，出現率に基づき関係値を算出した(式1)．$nvrelation(x,y)$は，概念xと概念yの間の主部・述部(NV: Noun Phrase and VerbPhrase)関係値，$nvfrequency(x,y)$は，概念xと概念yが，主部・述部関係で登場する頻度，$nvpair$は，テキストコーパス中に存在する主部・述部関係にある概念対を示す．

$$nvrelation(x,y) = \frac{nvfrequency(x,y)}{\sum nvpair} \quad (1)$$

連絡先: 森田武史，静岡大学大学院情報学研究科，〒432-8011 静岡県浜松市城北3-5-1，Tel: 053-478-1478，FAX 053-478-1478，e-mail: morita@ks.cs.inf.shizuoka.ac.jp

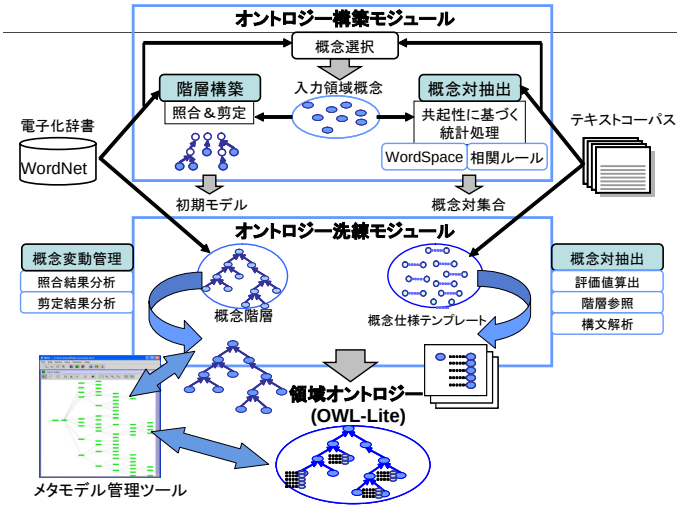


図 1: DODDLE-R システムフロー

概念階層の利用

テキストコーパス中には，オントロジー構築モジュールにおいて同定された入力概念の他にも，これに準ずる概念が存在する．例を図 2 に示す．テキストコーパス中に “seller” という概念が存在し，入力概念に “person” が存在した場合，電子化辞書を参照することにより，一般的に “seller” は “person” の特殊化概念であることがわかる．“buyer” についても同様のことがいえる．

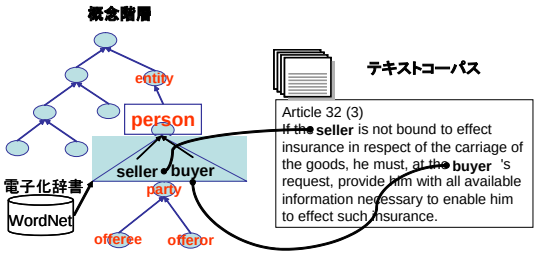


図 2: 文中の概念の入力概念への抽象化における例

図 3 では，概念階層中に抽象・具象関係として位置付けられた入力概念 A，B がある．文中の概念 X は，電子化辞書中で入力概念 A，B 両方を上位概念として持つ．この場合，概念階層中でより具象的な概念である入力概念 B を概念 X の抽象概念とする．

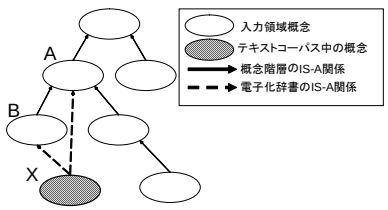


図 3: 上位概念が複数ある場合における文中の概念の入力概念への抽象化

以上の手法をテキストコーパスに対して適用し，入力概念への抽象化を行った後に，共起性に基づく統計処理と構文情報の抽出を再び行い，オントロジー構築モジュールで得られた概念対に対する重みづけの指標として用いる．

2.3 領域オントロジー構築支援環境とメタモデル管理ツールの統合

セマンティック Web における領域オントロジーの質を高めるために，領域オントロジー構築支援環境とメタモデル管理ツールの統合を行う．メタモデル管理ツールは，RDF と RDFS 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．メタモデル管理ツールのプラグイン機能を用いて， DODDLE-R 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メタモデル管理ツ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．メタモデル管理ツールのクラスエディタにオントロジー構築モジュールで構築された概念階層の初期モデルを入力し，ユーザは DODDLE-R が示唆する概念変動を行う箇所の編集を行う．２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-R 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＋概念定義）のバランスを見ながらバグの発見を行い，オントロジーの質を向上させる．

2.4 OWL 形式への変換とその課題

DODDLE-R によって構築される領域オントロジーは，概念階層と概念定義から構成される．概念階層は OWL が提供する `owl:Class` 要素及び `rdfs:subClassOf` プロパティによって定義する．概念定義は，概念対の間の関係を OWL におけるプロパティ，概念対をプロパティの定義域および値域としてとらえられる．概念定義は OWL が提供する `owl:ObjectProperty` 要素，`rdfs:domain` および `rdfs:range` プロパティによって定義する．

OWL におけるプロパティの定義域および値域の値はクラスでなければならないが，概念対の組み合わせにはクラスとクラス (C-C)，クラスとインスタンス (C-I)，インスタンスとインスタンス (I-I) の 3 通りが考えられる．4. 節で，インスタンスが概念対に含まれる可能性を示す．概念対のどちらか一方，または両方にインスタンスがあらわれる場合には，インスタンスの属するクラスをユーザが定義し，定義域または値域と同定する必要がある．概念対における概念のクラス及びインスタンスの識別は，領域オントロジーを OWL 形式でエクスポートするために必要な機能と考えられる．メタモデル管理ツールによりオントロジーを外在化する際にも必要である．

概念間の関係にはインスタンス間の関係をあらわすプロパティとクラス間の関係をあらわすメタプロパティの 2 種類が考えられる．図 4 にプロパティとメタプロパティを示す．クラス A とクラス B が存在するとする．プロパティはクラス A のあるインスタンスとクラス B のあるインスタンスの間に関係があることをあらわし，メタプロパティはクラス A のすべてのインスタンスとクラス B のすべてのインスタンスの間に関係があることをあらわす．両者の区別を行う機構も領域オントロジーを OWL 形式でエクスポートするために必要である．

3. DODDLE-R の実装

図 5 は，DODDLE-R のユーザインタフェースを示している．DODDLE-R は JAVA 言語で実装されている．DODDLE-R のユーザインタフェースは，Input Concept View, Taxonomic

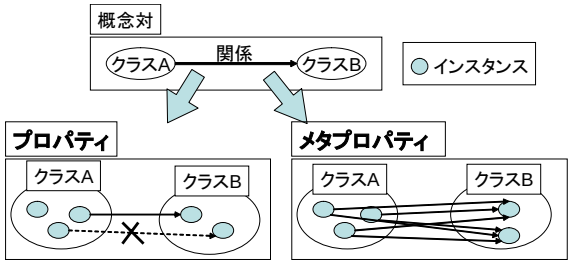


図 4: プロパティとメタプロパティ

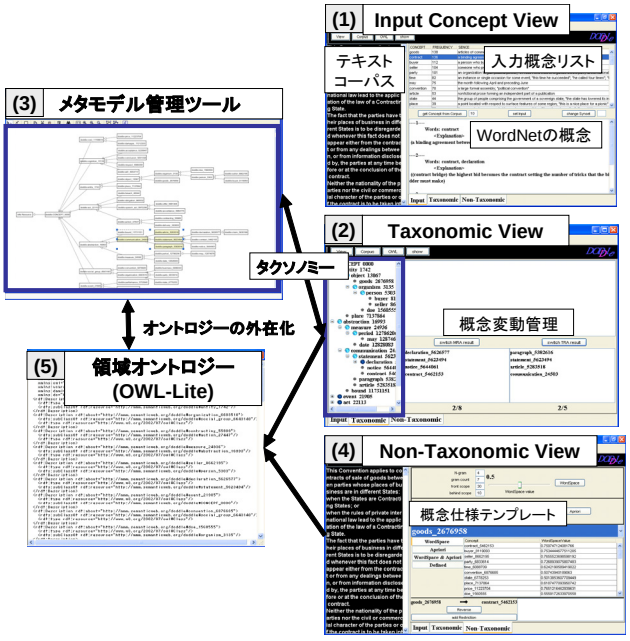


図 5: DODDLE-R のユーザインタフェース

View, Non-Taxonomic View の 3 つの部分から構成される．Input Concept View は図 1 のシステムフローにおけるオントロジー構築モジュールに対応する．Taxonomic View 及び Non-Taxonomic View は，図 1 のシステムフローにおけるオントロジー洗練モジュールに対応する．

Input Concept View (図 5 の (1)) では，テキストコーパスを入力として入力概念の候補となるリストを表示し，ユーザはそのリストの中から入力概念の選択を行う．入力概念の選択をする際には，Wordnet を参照し，入力概念と WordNet の概念との対応付けを行う．入力概念の決定後，概念階層の初期モデル及び概念対集合が得られる．

Taxonomic View (図 5 の (2)) では，メタモデル管理ツールとの連携により概念変動管理を行い，初期概念階層の洗練を行う．図 5 の (3) は，DODDLE-R が構築した初期概念階層をメタモデル管理ツールに読み込ませた時のスクリーンショットを示している．DODDLE-R が修正すべき初期概念階層内の概念集合を示唆し，ユーザはメタモデル管理ツールを通して修正箇所の編集を行う．また，洗練をおこなった概念階層を DODDLE-R にインポートする．

Non-taxonomic View(図 5 の (4)) では，概念定義を行う．ユーザは WordSpace および相関ルールにおけるパラメータの設定，構文情報および概念階層の利用による重み付けを行い，概念対集合の洗練を行う．概念対集合の中からユーザは重要と考えられる概念対を選択し，概念対の間の関係を補完し，概念定義を行う．

最終的に概念階層と概念定義を合成して，OWL 形式で領域オントロジーを得る (図 5 の (5))．構築された領域オントロジーはメタモデル管理ツールによって外在化され，ユーザはオントロジー全体のバランスを見ながらバグの発見を行う．

4. ケーススタディ

ケーススタディは概念定義における概念対の組み合わせとして，C-C，C-I，I-I の組み合わせが生じる可能性があることを示すために行った．クリエイティブ・コモンズ [CC][KANZAKI] が提供している 11 ライセンス文書のうちの 1 つ (Attribution-NoDerivs-NonCommercial) を DODDLE-R の入力として概念定義を構築する．クリエイティブ・コモンズのライセンス文書を選んだ理由は，ライセンス文書を機械処理可能なメタデータとして定義するための RDF スキーマ及び各ライセンス文書に対応するメタデータが提供されているからである．表 1 および表 2 にクリエイティブ・コモンズが提供する RDF スキーマ及びそのインスタンスの一覧を示す．

表 1: クリエイティブ・コモンズが提供する RDF スキーマにおけるクラス及びそのインスタンス

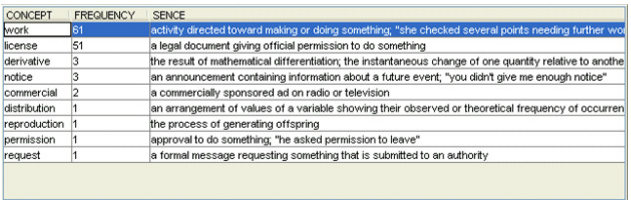
クラス	インスタンス
Work	
Agent	
License	PublicDomain
Permission	Reproduction, Distribution, DerivativeWorks
Requirement	Notice, Attribution, ShareAlike, SourceCode
Prohibition	CommercialUse

表 2: クリエイティブ・コモンズが提供する RDF スキーマにおけるプロパティ及びその定義域と値域

プロパティ	定義域	値域
license	Work	License
permits	License	Permission
requires	License	Requirement
prohibits	License	Prohibition
derivativeWork	Work	Work

ライセンス文書中の出現頻度に基づいて抽出した概念の中から，クリエイティブ・コモンズが提供するクラス及びそのインスタンスに類似する語彙を DODDLE-R の入力概念とする．図 6 に入力概念として選択した概念を示す．WordSpace により概念対を抽出した際の DODDLE-R のスクリーンショットを図 7 に示す．図 7 の上部は，WordSpace 及び相関ルールにおけるパラメータの設定画面を示している．図 7 の下部は，WordSpace により抽出された，概念 work と対になる概念のリストを示している．最終的に選択された概念対を表 3 に示す．

表 3 より，11 概念対が抽出され，C-C，C-I，I-I の 3 通りの組み合わせが生じることがわかった．表 3 中の概念対のう



CONCEPT	FREQUENCY	SENCE
work	61	activity directed toward making or doing something; "she checked several points needing further wo
license	51	a legal document giving official permission to do something
derivative	3	the result of mathematical differentiation; the instantaneous change of one quantity relative to anothe
notice	3	an announcement containing information about a future event; "you didn't give me enough notice"
commercial	2	a commercially sponsored ad on radio or television
distribution	1	an arrangement of values of a variable showing their observed or theoretical frequency of occurren
reproduction	1	the process of generating offspring
permission	1	approval to do something; "he asked permission to leave"
request	1	a formal message requesting something that is submitted to an authority

図 6: ライセンス文書から抽出された入力概念

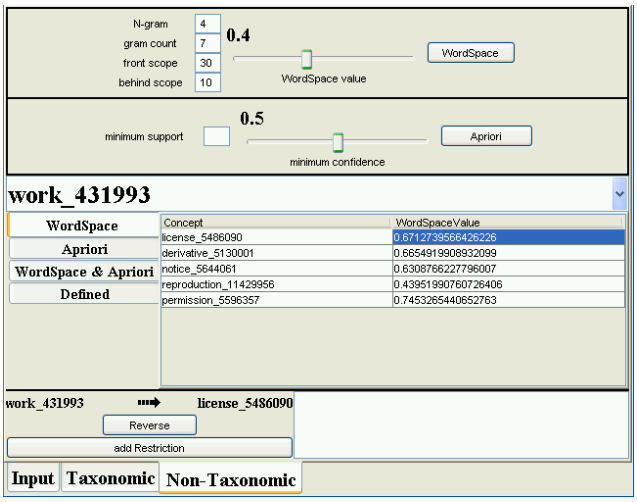


図 7: WordSpace により概念対を抽出した際の DODDLE-R のスクリーンショット

ち，クリエイティブ・コモンズが提供する RDF スキーマにおけるプロパティの定義域および値域と一致するものには下線を引いてある．定義域および値域と一致した 4 概念対中 2 概念対は C-I の関係にあり，インスタンスの属するクラスを定義することによって，定義域および値域を同定できる．表 1 より `derivative( DerivativeWorks)` は `Permission` クラスのインスタンスであることから，`license-derivative` は `license-permission` とみなすことができる．同様に `license-notice` は `license-requirement` とみなすことができる．以上より，表 2 および表 3 の範囲内における抽出概念対の正答率は 3/10，再現率は 3/5 である．

本ケーススタディではクラス及びインスタンスの区別が明確な状態で入力概念を選択したため，インスタンスが概念対に含まれることは予想できた．実際には入力概念を選択する時点で，入力概念がクラスかインスタンスかを区別することは難しい．領域オントロジーを OWL 形式でエクスポートするためには， DODDLE-R のオントロジー洗練モジュールにクラスかインスタンスかを同定するための機構が必要である．

5. おわりに

本稿ではセマンティック Web のためのオントロジー構築支援環境 DODDLE-R について述べた．共起性に基づく統計処理に加えて，構文情報と階層情報を利用した概念対に対する重み付けの指標を提案した．領域オントロジーを OWL 形式でエクスポートする際に，概念対における概念がクラスかインスタンスかを同定する必要があることをケーススタディを通じて明らかにした．概念間の関係としてプロパティとメタプロパティを区別する必要があることも明らかとなった．OWL エクスポート時における，クラス・インスタンス，プロパティ・メタプロパティ同定の支援手法の開発は今後の課題である．

表 3: WordSpace により抽出された概念対の分類 (パラメータ: N-gram: 4, gram count: 7, front scope: 30, behind scope: 10, value: 0.4)

C-C	C-I (I-C)	I-I
work-license	work-derivative	derivative-notice
work-permission	work-notice	
license-permission	work-reproduction	
	license-derivative	
	license-notice	
	notice-permission	
	commercial-request	

謝辞

本研究は，文部科学省科学研究費補助金 (15300043)「オントロジーとモデル駆動型アーキテクチャの統合」の助成によるものである．

参考文献

[Kurematsu 04] Masaki Kurematsu, Takamasa Iwade, Naomi Nakaya, and Takahira Yamaguchi, DODDLE II: A Domain Ontology Development Environment Using a MRD and Text Corpus, IEICE TRANS. INF. & SYST., VOL.E87-D, NO4, 2004

[Morita 03] Takeshi Morita, Noriaki Izumi, Naoki Fukuta and Takahira Yamaguchi, Meta-Model Management Environment for RDF Contents, 2nd International Conference on Global Research and Education(Inter-Academia 2003), PROCEEDINGS VOLUME 2, pp. 307-314, 2003.

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

[R.Agrawal 94] Rakesh Agrawal and Ramakrishnan Srikant, "Fast algorithms for mining association rules", Proc. of VLDB Conference, pp.487–499, 1994,

[RDF] Ora Lassila and Ralph R. Swick, "Resource Description Framework(RDF) Model and Syntax Specification", http://www.w3.org/RDF/, 1999,

[OWL] Michael K. Smith, Chris Welty and Deborah L. McGuinness, "OWL Web Ontology Language Guide", http://www.w3.org/TR/owl-guide/,

[CC] Creative Commons, http://creativecommons.org/,

[KANZAKI] クリエイティブ・コモンズのメタデータ, http://www.kanzaki.com/docs/sw/ccm.html