Vol.2012-MPS-91 No.32
Vol.2012-BIO-32 No.32
2012/12/7

書籍オントロジーのための推論部の開発

岩片 悠里[1]　川嵜 美波[1]　江口 由記[1]　高田 雅美[2]　城 和貴[2]

概要：本稿では，教科書を基に作られたオントロジーから情報を得るための推論部の開発を行う．推論エンジンは既存のものを用い，推論規則を新たに定義する．この推論部では，ユーザから与えられた問題文から，答えの導出を行う．これにより，ユーザの負担を軽減させることを可能にする．さらに，専門書の知識を基に推論の知識モデルとなる書籍オントロジーを作成することによって，効率的な情報の取得を目指す．

キーワード：オントロジー，推論，推論エンジン，Jena，人工知能，セマンテイック Web

Development of Inference Part for Book Ontology

Abstract: In this paper, we develop inference part for an ontology that is based on the textbook. We use an existing inference engine and define new inference rules. By using this inference part, the answer can be derived from a problem statement. Thus, it is possible to reduce the burden on users. The book ontology is construted based on the textbook as knowledge model of the inference. Therefore, it is possible to retrieve information efficiently.

Keywords: Ontology, Inference, Inference engine, Jena, Artificial intelligence, Semantic Web

1. はじめに

現代社会では，Web 技術の発展により，誰でも気軽に情報を発信できるようになっているため，Web 上の情報量が爆発的に増加しており，雑多な情報も多い．そのため，得たい情報が他の膨大な情報にまぎれてしまい，正確な情報を取得する手間がかかってしまう．そこで，より効率的に情報を取得するための方法が求められている．

その方法の一つにエージェントによる自律的な推論がある．例えば，Web 上で検索を行う場合，必要とする情報に含まれていると考えられるキーワードを予測し，検索ワードとして設定する必要がある．必要とする情報が見つからない場合は，別の検索ワードを与えなければならない．これらの繰り返しは非効率的である．一方，推論によって情報を取得する場合，ユーザが必要とする情報を，計算機が与えてくれるので，簡単に欲しい情報にたどり着くことができる．結果として，情報取得に要する手間は大幅に軽減されると考えられる．

本稿では，オントロジーを用いた推論を行う．オントロジーとは，計算機が情報を理解できるように，基本的な概念の定義や概念同士の関係性を付加したもののことを指す．これを用いることで，人間の理解している対象世界を体系的にモデル化し，計算機に与えることができる．オントロジーを用いた推論では，インスタンスモデル，インスタンスをあてはめるための知識モデル，及び推論規則から推論モデルを作成し，この推論モデルから新たな知識を導き出す．インスタンスモデルは，個別の概念を知識体系化したものである．知識モデルは，インスタンスモデルをあてはめるための知識を体系化したものである．推論規則は，推論を行うための規則を設定したものである．知識モデルを作成するためには，モデルとなる知識のリソースが必要となる．しかし，そのリソースを示す情報源の信ぴょう性が低いと推論結果も正確なものが得られないかもしれない．そこで本稿では，専門書の内容を基に知識モデルを構築することにより，信ぴょう性の高い情報の取得を目指す．

専門書を基に作られたオントロジーから情報を得るために推論規則を新たに定義し，推論についての評価を行う．

[1] 奈良女子大学　理学部　情報科学科
Department of Information and Computer Sciences, Nara Women's University, Nara, Nara 630–8506, Japan
[2] 奈良女子大学 研究院 自然科学系 情報科学領域
Academic Group of Information and Computer Sciences, Nara Women ' s University, Nara, Nara 630–8506, Japan

Vol.2012-MPS-91 No.32
Vol.2012-BIO-32 No.32
2012/12/7

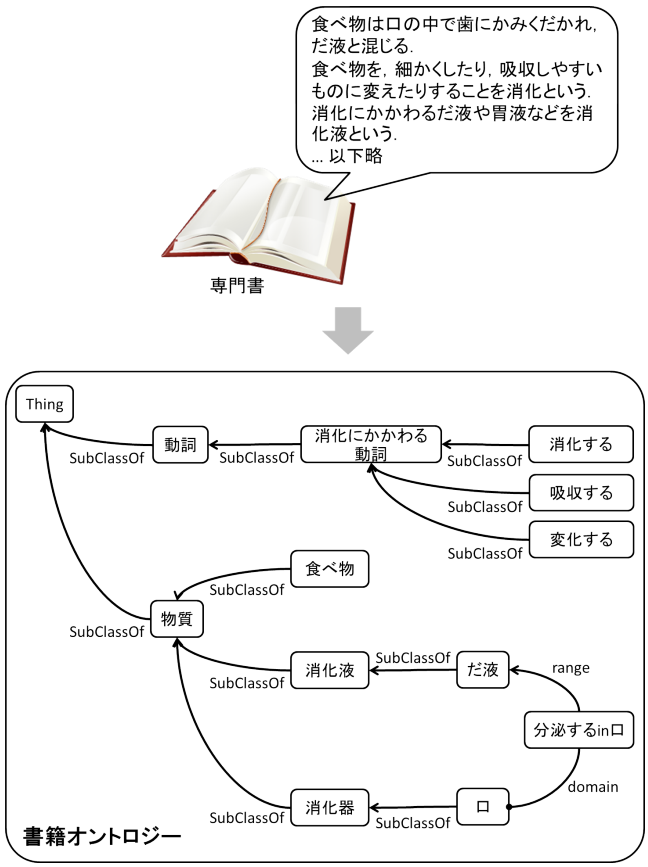

図 1 書籍オントロジーの概要図

ここで使用する推論エンジンには，既存のものを用いる．利用するオントロジーのリソースには，最も簡単な専門書としてあげられる教科書の内容を扱う．その際使用した教科書に準拠している問題集の問題を用いて評価実験を行い，有効性を調べる．

本稿では，2 章で書籍オントロジーについて述べる．3 章では，書籍オントロジーのための推論部についての説明を行う．4 章では，評価実験について述べる．

2. 書籍オントロジー

計算機が質問文からユーザが必要とする情報を推測するためには，質問文の意味を理解する必要がある．そこで，オントロジーが必要となる．

本章では，質問文の意味を理解するために，オントロジーを利用して，質問文の答えを導く際に必要な知識の体系化を行う．知識には，専門書の内容を用いる．このモデルを書籍オントロジーとする．**図 1** は，書籍オントロジーの概要図を示したものである．書籍オントロジーは，クラスを階層関係になるように繋いだものである．クラスは，書籍の本文から，主体となる動詞と名詞を抜き出し，その中から非自立語や代名詞を抜き出したものと，それらを分類するものから作成する．加えて，各クラス同士を結び付けるためのプロパティがある．プロパティは，クラス同士の関係性を表すものである．

本稿では，最も簡単な専門書として教科書を採用する．さらに，分野を理科の「食べ物の消化と吸収」に限定する．理科の「食べ物の消化と吸収」は，概念とそれに所属するインスタンスの関係が明確なため，推論規則を定義しやすい．専門書として教科書を用いる場合，次のような点に注意しなければならない．

クラスに関して注意しなければならないことは 2 つに分けられる．1 つ目は，クラスとする語句を選ぶ際は，分野に即した語句を中心に選定することである．本稿では，理科の教科書の「食べ物の消化と吸収」に分野を限定している．そのため，消化を行う器官や消化を行うもの，消化されるもの，吸収を行う器官や吸収されるものなど，消化や吸収に直接関係していることが読み取れる動詞や名詞を中心にクラスを作成している．また，それらの動詞や名詞を分類するための語句もクラスとして作成している．2 つ目は，本文に記載されていない事柄も，適宜，クラスとして追加することである．特に小学生を対象とする教科書では，教科書には，イラストや写真が多い．言葉では記述せずに，それらのイラストや写真から理解を促す事柄も少なからず存在する．また，イラストや写真の下に，重要な語句がイラストや写真の名称として，説明もなしに小さく載っている場合もある．そのため，イラストや写真など本文に記載されていない事柄も，適宜，クラスとして追加する必要がある．

プロパティに関して注意しなければならないことは，必要最小限のクラス同士の関係を表すために付加すべきであることである．問題を解くうえで必要となる語句は，教科書に記述されている動詞や名詞の全てではない．さらに，語句同士の関係が必要となるものは，限られている．オントロジーの上位階層を参照することで，事足りるものも多い．結果として，全てのクラス間にプロパティを付与する必要ない．

図 2 は，書籍オントロジーを Protégé[5] の機能を用いて可視化したものである．ただし，見易さのため，一部を省略している．長方形は，クラスを表している．紫色の線で結ばれているクラスは，subClassOf 関係によって結ばれているクラス同士であることを表している．紫色以外の線で結ばれているクラスは，プロパティでクラスが結ばれているクラス同士であることを表している．以下に書籍オントロジーのクラス及びプロパティを示す．

- クラス
  - 物質
    - 液体
      - 消化液
    - 消化器
    - 食べ物
      - でんぷんを含む食べ物

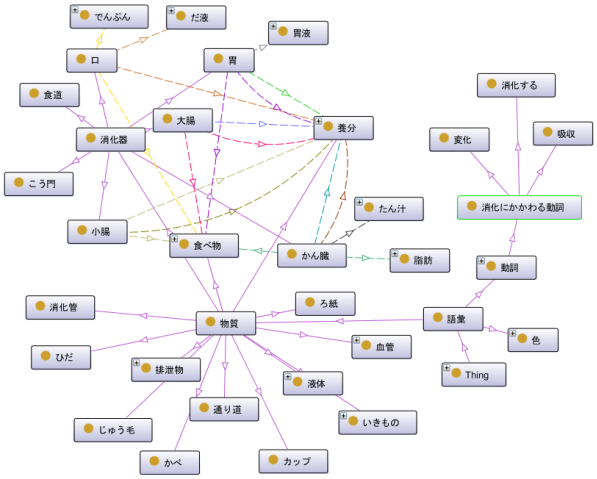

図 2 書籍オントロジーの可視化

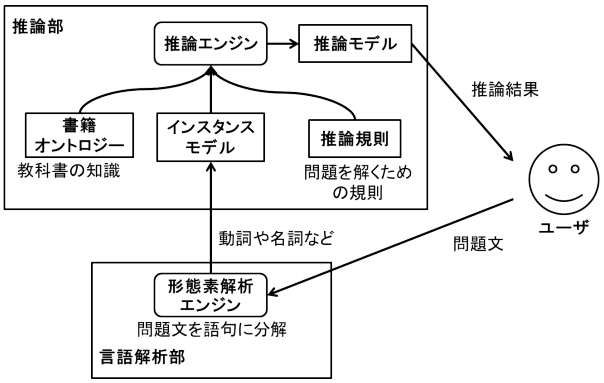

図 3 言語解析部を加えたシステムの概要

  - 養分
  - いきもの
  - 血管
- 動詞
  - 消化にかかわる動詞
- 色

- プロパティ
  - 消化する
  - 吸収する
  - 分泌する
  - 蓄える

3. 書籍オントロジーのための推論部

推論を行うためには，インスタンスモデル，書籍オントロジー，推論規則から推論モデルを作成する必要がある．本章では，これらの作成，及び設定方法について述べる．

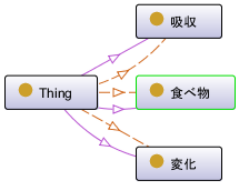

図 4 インスタンスモデルの可視化

3.1 概要

本推論部では，入力された問題文に対して，推論結果として答えを返す．推論部では，言語解析部が問題文から取得した名詞や動詞からインスタンスモデルを作成し，推論を行う．**図 3** は，本推論部に言語解析部を加えたシステムの概要図を示したものである．

ユーザによって入力された問題文は，言語解析部に送られる．言語解析部では，既存の形態素解析エンジンを用いて，問題文から動詞と名詞を取得し，これらを推論部に送る．ここで用いるオントロジーは，2 章で説明したものである．推論部では，書籍オントロジーを参照して，インスタンスモデルの自動作成を行う．作成されたインスタンスモデル，書籍オントロジー，及び推論規則から，既存の推論エンジンを用いて推論モデルを自動にて作成する．最後に，推論モデルから得られる推論結果を問題の答えとしてユーザに返す．

3.2 推論部の手順

本節では，推論部の処理手順について述べる．推論部では，以下の手順の処理を行っている．

(1) 空のインスタンスモデルを作成
(2) 語句と同一クラスの取得
(3) 問題文に記述されていることを示すプロパティを取得
(4) クラスとプロパティを結合し，インスタンスモデルに追加
(5) 推論モデルの作成
(6) 推論結果の取得

手順 (1) では，問題文に関連するクラスやプロパティを追加し，必要最小限のモデルから効率良く推論を行うため，空のインスタンスモデルを作成する．この時点で，インスタンスモデル内には，クラスやプロパティは存在しない．

手順 (2) では，書籍オントロジーの定義に添わせるため，書籍オントロジーの中から言語解析部が取得している語句と同一の語句が label 関係で結ばれているクラスを取得する．さらに，同じクラスをインスタンスモデル内に作成する．これにより，推論モデルを作成する際に，同じ事柄を指すクラスが異なるクラスとして認識されることを防ぐ．

手順 (3) では，手順 (2) と同様に，書籍オントロジー

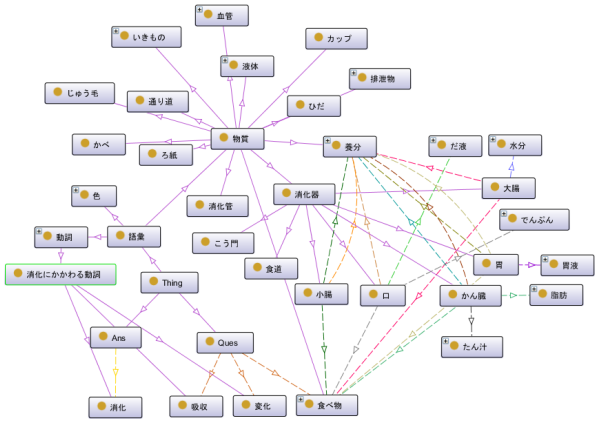

図 5 推論モデルの可視化

の定義に添わせるため，書籍オントロジーの中から問題文に記述されていることを示すプロパティを取得する．さらに，同じプロパティをインスタンスモデル内に作成する．理由は，手順（2）と同じである．

手順（4）では，手順（2）及び手順（3）で取得しているクラスとプロパティを結びつける．これらを結合させることによって，クラスが，問題文に記述されているクラスであることを表すことが可能となる．さらに，結合させたものをインスタンスモデルに追加する．**図 4** は，「食べ物を吸収しやすいものに変えるはたらきを、何と言いますか？」という問題から作成されたインスタンスモデルを，Protégé[5] の機能を用いて可視化したものである．長方形で囲まれている「食べ物」,「吸収」,「変化」が，インスタンスモデル内に作成されているクラスを表す．「Thing」は，クラスの根となるクラスである．オレンジ色の線が，手順（3）で取得している問題文に記述されていることを示すプロパティで結ばれていることを表す．紫色の線は，クラス同士が subClassOf 関係を表す．

手順（5）では，推論を行うための推論モデルの作成を行う．推論モデルは，インスタンスモデル，書籍オントロジー，及び推論規則から作成される．推論モデルの作成には既存の推論エンジンである Jena[7] を用いる．**図 5** は，「食べ物を吸収しやすいものに変えるはたらきを、何と言いますか？」という問題から作成された推論モデルを，Protégé[5] の機能を用いて可視化したものである．下方にある「Ques」が問題文に記述されているクラスを結ぶためのクラスである．同じく下方にある「Ans」は，問題の答えとなるクラスを結ぶためのクラスである．図 5 から，「Ques」には，「食べ物」,「吸収」,「変化」がオレンジ色の線で表されているプロパティによって結び付けられていることがわかる．また，「Ans」には，「消化」が黄色の線で表されるプロパティで結び付けられている．これは，「消化」が問題の答えとなるクラスであることを示す．

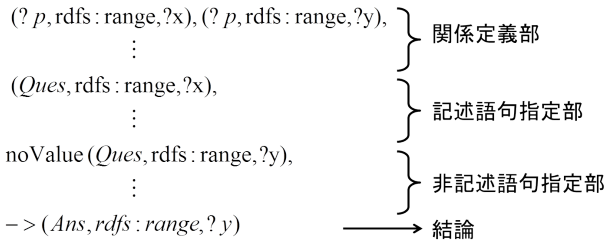

図 6 推論規則の構成

手順（6）では，推論モデルから推論結果を取得する．推論結果は，問題の答えを指すクラスをあらかじめ設定しておき，これを参照することで導き出される．

3.3 推論規則

本章では，推論規則の設定方法について述べる．本稿において，推論規則は，全て前向き推論の形式をとっている．前向き推論の形式とは，$(node, node, node)\ -> (node, node, node)$ のように矢印が右向きになっている推論規則のことである．矢印より左辺が前提であり，右辺が結論である．インスタンスモデル，及び書籍オントロジーと照らし合わせて，この前提が満たされていれば，推論規則に従い，結論を実行する．この結論実行が開始されることを，発火という．$(node, node, node)$ はトリプルパターンを指す．このトリプルの中身は，主に教科書の本文を基に定義している．

推論規則は，次のような法則で定義している．教科書の本文から，主体となる名詞を取り出し，それを解とする．さらにその周辺にある動詞と名詞を取り出し，その中から非自立語や代名詞を抜き出す．残りの語句から，解を導き出すための語句を最少個数となるように絞る．それらの語句が，問題文に含まれていることを推論規則の前提とし，最初に取り出してある解を答えとすることを推論規則の結論となるように推論規則を定義していく．

図 6 は推論規則の構成を表している．$?x$，$?y$，$?z$ は，変数を示す．推論規則の前提は，関係定義部，記述語句指定部，非記述語句指定部に分かれる．関係定義部では，次のようなトリプルパターンを使用して，変数の関係を定義している．

$$(?p, rdfs : domain, ?x), (?p, rdfs : range, ?y)$$

$$(?p, rdfs : subClassOf, ?x)$$

$$(?p, rdfs : subPropertyOf, ?x)$$

$rdfs : domain$ や $rdfs : range$，$rdfs : subClassOf$，$rdfs : subPropertyOf$ は，RDF スキーマの用語を示す．1

つ目の式は，?x と ?p が domain 関係で結ばれていること，?y と ?p が range 関係で結ばれていることを表す．これを前提の要素とし，これらの変数を用いて他の要素を記述することにより，クラス同士の関係性を推論規則に記述することが可能となる．

記述語句指定部では，問題文に記述されている語句を指定する．$Ques$ は，問題文に記述されている語句を示すクラスを指すためのプロパティである．

非記述語句指定部では，問題文に記述されていない語句を指定する．主に，結論部で答えとして指定するクラスをこの要素で指定する．

結論には，Ans が答えとして指定するクラスを指すように設定する．Ans は，答えとして指定するクラスを指すプロパティである．

以上のように，推論規則を設定する．ただし，教科書には，イラストや写真も多く，言葉では，記述していないことも多い．また，イラストの説明文に重要な単語が，説明文もなしに小さく載っている場合もあるため，適宜，本文を補い，同様の方法で推論規則を設定する必要がある．

4. 評価

本章では，推論部の有効性を調べるための評価実験について述べる．評価実験では，問題文を入力し，出力された推論結果が問題の答えと一致するかどうかを調べる．問題には，書籍オントロジーの構築の際に使用した教科書に準拠している問題集の問題と一般に販売されている問題集の問題を用いる．本稿では，分野を「消化と吸収」に限定しているため，評価実験で使用する問題も「消化と吸収」の分野に関する問題に限る．また，問題形式が多様であるため，入力する問題文は以下ように制限を与える．

- イラストを見て答える問題や 2 択問題は，除去
- 穴埋め問題は，空欄に「何」の文字を挿入
- 空欄が複数ある穴埋め問題は，空欄が 1 つになるように修正

この制限を満たす問題は，教科書ワーク理科から 10 問，小学/理科まとめ上手から 15 問の合計 25 問である．これらの問題を用いて評価実験を行う．

書籍オントロジーは，Protégé[5] を用いて，手動で構築を行う．手動構築は，自動構築に比べ，信頼性が高い．評価実験では，啓林館 [8] が出版している小学 6 年生の理科の教科書であるわくわく理科 6 を基にして書籍オントロジーの構築を行う．インスタンスモデルの作成及び，推論モデルの作成には Jena[7] を使用している．Jena[7] は，セマンティック Web のための Java フレームワークである．教科書に準拠している問題集には，教科書と同じく啓林館 [8] が出版している教科書ワーク理科を用いる．一般に販売されている問題集には，受験研究社 [9] が出版している小学/理科まとめ上手を用いる．

表 1 評価実験の結果

	正答数	誤答数	正答率
教科書ワーク理科	9	1	90%
小学/理科まとめ上手	3	12	20%

評価実験の結果は，**表 1** のとおりである．1 列目は，問題の記載されていた問題集を示す．2 列目は，正しい答えを導いた回数を正答数として示す．3 列目は，正しくない答えを導いた回数を誤答数として示す．4 列目は，100 (正答数) / (正答数 + 誤答数) を正答率として示す．

表 1 より，教科書ワーク理科の正答率は，90%という高い数字が出ている．一方，小学/理科まとめ上手は 20%である．正しい答えを導かない場合は，次の 3 つに分けられる．

1 つ目は，何も出力されない場合である．小学/理科まとめ上手の 10 問がこれに分類される．何も出力されない場合は，次の 4 種類のことが原因で起こると考えられる．1 種類目の原因は，推論規則の設定漏れのためである．小学/理科まとめ上手の 3 問がこれに分類される．推論規則の設定には，教科書から答えとなり得る名詞を選択する際に，設定者によって個人差が生じる．そのため，その選択を誤ると，推論規則の設定漏れにつながる．また，教科書には，言葉では記述せずに，イラストや写真から，理解を促す事柄が多い．そのため，文字だけで記載されている書籍よりも，推論規則の設定漏れが起こりやすい．この場合は，新しく推論規則を設定することで解決できると考えられる．2 種類目の原因は，前提が厳しすぎるため，発火しない場合である．小学/理科まとめ上手の 1 問がこれに分類される．この場合は，発火しない原因となる要素を前提から削除することで解決できると考えられる．削除対象となる前提の要素は，記述語句指定部の要素とする．ただし，適切に要素を削除を行わなければ，他の問題で正しい答えを導くことができない恐れがあるため注意が必要である．3 種類目の原因は，言語解析部の語句の取得に誤りがあるために，インスタンスを取得できない場合である．小学/理科まとめ上手の 1 問がこれに分類される．評価実験では，「じゅう毛」を「じゅう」と「毛」に分けて取得している．この場合は，言語解析部で「じゅう」の直後に「毛」を取得する際，これらを結合させて「じゅう毛」に修正して推論部に送ることで解決できると考えられる．ただし，異なる意味を表すものであっても自動的に語句同士を結合してしまうため，注意が必要である．4 種類目の原因は，問題を解く際に，教科書にはない知識を必要としているため，書籍オントロジーの知識では対応できない場合である．小学/理科まとめ上手の 5 問がこれに分類される．一般に販売されている問題集では，受験用に教科書にない知識を問う問

題も載せてある．そのため，本推論部には，その問題に対応する推論規則が存在しない．

2つ目は，異なる答えを出力する場合である．小学/理科まとめ上手の2問がこれに分類される．異なる答えを出力する場合は，次の2種類のことが原因で起こると考えられる．1種類目の原因は，前提が厳しすぎるため，該当する推論規則に発火しないうえに，他の推論規則に発火する場合である．小学/理科まとめ上手の1問がこれに分類される．この場合は，該当する推論規則の前提から発火しない原因となる要素を削除し，発火する他の推論規則の前提に要素を追加することで解決できると考えられる．削除対象となる前提の要素は，記述語句指定部の要素とする．また，要素の追加は，非記述語句指定部の要素とする．ただし，記述語句指定部の要素を削除する際は，何も出力されない場合の2種類目の原因と同様の理由により，他の問題に対する推論に影響を与えないようにしなければならない．また，非記述語句指定部の要素の追加は，問題文中に含まれる語句次第で，発火しなくなる推論規則が増えることを意味する．ゆえに，適切に要素の追加を行わなければ，今まで正しい答えを導いていた問題まで答えを導くことができなくなる恐れがあるため注意が必要である．2種類目の原因は，問題を解くためには，教科書にない知識を必要としているため，書籍オントロジーの知識では対応できない場合である．小学/理科まとめ上手の1問がこれに分類される．教科書にない知識であるため，その知識のための推論規則が設定されない．また，他の問題のために設定している推論規則が発火する可能性がある．そのため，この問題を解くことができない．

3つ目は，複数の答えを出力する場合である．教科書ワーク理科の1問がこれに分類される．この原因は，該当する推論規則以外に，他の推論規則に発火するため，複数の答えが導き出されると考えられる．この場合は，発火する他の推論規則の前提に要素を追加することで解決できると考えられる．また，要素の追加は，非記述語句指定部の要素とする．ただし，非記述語句指定部の要素の追加は，推論規則が発火するための条件を厳しくすると同義であるため，正しい答えを導いている問題に対しても推論規則の発火が行われなくなる可能性があるため，注意が必要である．

以上より，問題集によって正答率に差が発生しているのは，推論規則の設定や教科書に記述されていない知識を問う問題が含まれていることが原因であると考えられる．教科書ワーク理科は，教科書に準拠した問題集であるため，教科書の本文に則した形式で問題を出題している．ゆえに，推論規則の設定漏れが少ないうえに，教科書に記述されていない知識を問う問題が存在しない．そのため，小学/理科まとめ上手との正答率に差が生じたと考えられる．

5. まとめ

本稿では，教科書を基に作られたオントロジーから情報を得るための推論部の開発を行っている．推論部では，まず，問題文の語句からインスタンスモデルの作成を行う．次に，このインスタンスモデルと教科書をもとに構築した書籍オントロジー，新たに定義した推論規則から，推論モデルの作成を行う．この推論モデルから，推論結果を導出する．推論エンジンには，Jena[7] を用いている．評価実験では，教科書に準拠した問題集の問題を使用した場合は，90%の正答率，一般に販売されている問題集の問題を使用した場合は，20%の正答率を得ている．考察の結果，教科書に記述されていない知識を問う問題を除くと，推論規則を修正することにより，ほぼ全ての問題で正しい答えを導くことが可能と考えられる．

推論規則の設定漏れは，必ず起こりうる．また，書籍オントロジーの規模が拡張していけば，推論規則を手動で設定することにいずれ限界が生じる．これを踏まえて，今後は，正しい答えを導くことのできない問題から，新たな推論規則を導き，追加を行う機能を拡張させることを目指す．これにより，より広い分野の問題の答えを求めることが可能となる．

参考文献

[1] 和泉 諭, 加藤 靖, 高橋 薫, 菅沼 拓夫, 白鳥 則郎:オントロジを利用した健康支援システムの提案とその評価, 情報処理学会論文誌,Vol.49,No.2,pp.822-837(2008).
[2] 竹之内 隆夫, 岡本 直之, 川村 隆浩, 大須賀 昭彦, 前川 守:ユビキタス環境において動的なコンテクストに応じて知識情報をフィルタリングする推論エージェントの開発, 電子情報通信学会論文誌,D-I J88-D-I(9),pp.1428-1437(2005).
[3] 小林一郎：人工知能の基礎，サイエンス社 (2008).
[4] 溝口理一郎：オントロジー構築入門，オーム社 (2006).
[5] A WEB Page, the Stanford University School (online), available from ⟨http://protege.stanford.edu/⟩ (accessed 2012-11-06).
[6] A WEB Page,Kyoto University (online), available from ⟨http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html⟩ (accessed 2012-11-06).
[7] A WEB Page,HP Labs (online), available from ⟨http://jena.apache.org/index.html⟩ (accessed 2012-11-06).
[8] A WEB Page, 啓林館 (online), available from ⟨http://www.shinko-keirin.co.jp/⟩ (accessed 2012-11-06).
[9] A WEB Page, 受験研究社 (online), available from ⟨http://www.zoshindo.co.jp/⟩ (accessed 2012-11-06).