交通オントロジーを対象とした質問文の SPARQL クエリ変換

Converting Questions into SPARQL Queries for Traffic ontology

鈴木 遼司　　三輪 誠　　佐々木 裕

Ryoji Suzuki　Makoto Miwa　Yutaka Sasaki

豊田工業大学

Toyota Technological Institute

{sd10033, makoto-miwa, yutaka.sasaki}@toyota-ti.ac.jp

１．はじめに

現在, 交通に関する知識ベース構築が盛んに行われている. だが現状,ユーザが知識ベースに求める情報を得る方法として, 自然言語文ではなく, 問い合わせるための言語で記述しなければならない. 本研究ではその問題を解消するために, オントロジーで表現された知識から, 必要な情報を取得する際の自然言語で与えられた質問文を, 論理式を介して, 問い合わせ言語である SPARQL クエリ言語に変換することを目的とする.

2. 関連研究

2.1　Scaling Semantic Parsers with On-the-fly Ontology Matching

Kwiatkowski[1]らは, 大規模な知識ベースから必要な情報を抽出する際, 質問文を CCG[1] を用いて解析し, 論理式に変換する事を採用している. CCG は辞書といくつかの規則で構成されている. 辞書は多数の語彙項目を含んでおり, 各単語にカテゴリを割り当てる. CCG は質問文を単語に分け, 辞書や規則を用いて各々の単語に品詞と隣り合う単語との関係を示す. それらの情報を用いる事で論理式を構成している.

2.2 日本語係り受け解析ソフト Cabocha

係り受け解析とは, 文章を文節ごとに分け, 各文節が同文中のどの文節に係るのかということを解析する日本語係り受け解析器である Cabocha を用いている[2]. Cabocha の特徴としては,

1) Support Vector Machines(SVMs)に基づく高性能な係り受け解析器である.
2) IREX の定義による固有表現解析が可能であり, 平文はもちろん, 形態素解析済みデータ, 文節区切りデータ, 部分的に係り関係が付与されたデータからの解析が可能である.
3) 係り受けの同定に使用する素性をユーザー側で再定義可能である.
4) データを用意すれば, ユーザー側で学習を行うことが可能である.

といった特徴をもつ.

2.3 オントロジエディタ Protégé

Protégé[3]は公開サイトのデータによると, 24344 人 (2015 年 1 月現在) の登録ユーザーがあり, 名実ともに現時点では最もよく使われているオントロジエディタといえる. どちらかといえば, オントロジーの構築のフェーズよりもむしろ, 利用フェーズに重点が置かれており, オントロジーの構築後の知識獲得ツールとしての機能, オントロジーの併合/調整の機能に加えて, 利用者固有の機能拡張を可能にするプラグイン機能が備わっている. Protégé の主な特徴は以下の四つにまとめることが出来る.

(1) 利用者が基本表現を再定義して知識のモデルを拡張する機能
(2) オントロジーの出力形式を任意の形式言語にカスタマイズする機能
(3) インタフェースをカスタマイズする機能
(4) Reasoner 等の応用プログラムを組み込むための強力なプラグイン機能

これらの機能により，Protégé はドメインモデルを利用するためのメタツールとなっている．

2.4 問い合わせ言語 SPARQL

SPARQL[4]とは問い合わせ言語の一種であり，RDF データを検索するために広く使われている言語である．SELECT 句で RDF トリプル（主語，述語，目的語）のパターンを指定してそれにマッチするものを調べることで検索結果を変数に結び付けられた検索結果を取り出す方法の他，CONSTRUCT 句を用いて結果から新しい RDF データを生成する方法，DESCRIBE 句で結果を記述している RDF データを探して返す方法，ASK 句で検索パターンがマッチするかどうかだけを単純に返す方法がある．また，結果件数の上限を設ける LIMIT 句も用意されている．

3. 提案手法

本研究では日本語を対象とし，広く使われている日本語構文解析器の一つである Cabocha を使用し，質問文として

①下位概念数を問い合わせる質問文
②下位概念を問い合わせる質問文

を対象として質問応答を行う．知識ベースを構築し，質問文を SPARQL クエリに変換し，解答抽出システムを使う事で構築した知識ベースに問い合わせ，獲得した結果を解答とする．具体的な処理の流れを図 1 に示す．

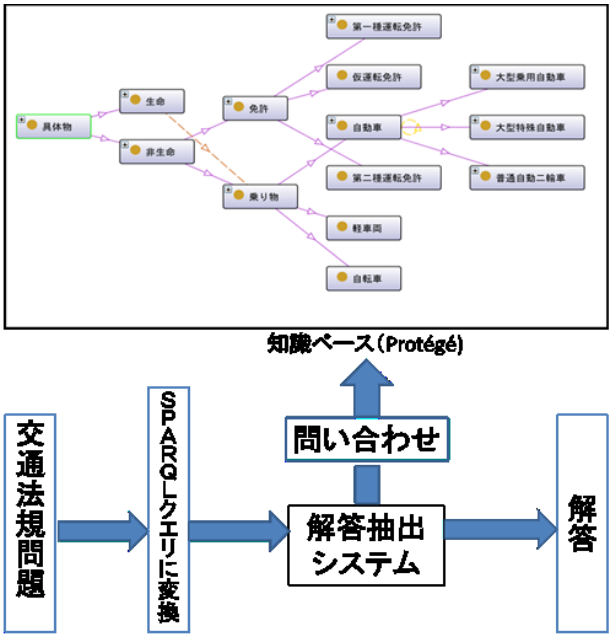

図 1. 処理の流れ

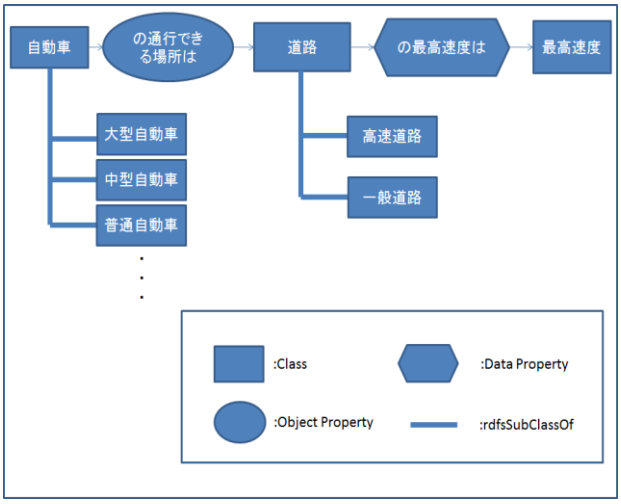

図 2. 概念関係図の一部の例

本節では，このうち，知識データベースの構築と自然言語文の SPARQL クエリ変換について説明する．

3.1 知識データベースの構築

オントロジーに問い合わせるためには自然言語から SPARQL クエリ言語に変換する必要があり，本研究では交通用語を対象とし，オントロジエディタ Protégé を用いて知識ベースを構築する．交通用語は自動車教習の教本や教則冊子等[5][6][7]から収集した．Class，Object Property，Data Property に分類し，概念関係

を体系化をする．例として概念関係図の一部を図 2 に示す．それぞれの定義は以下の通りである．

・Class⇒モノやコトのクラスを表す概念

・Object Property⇒クラス間の関係を示す概念

・Data Property⇒モノやコトの属性を示す概念

モノやコトの上位・下位概念を分類し，データベースを構築する．

3.2 質問文の SPARQL クエリへの変換

自然言語質問文から SPARQL クエリに変換する過程においてのステップは以下の 3 つからなる．

(1) 主語，動詞，目的語で構成される質問文に対して構文解析を行い品詞・係り受け関係を獲得し図 3 のようなフォーマットに照らし合わせる．

(2) 質問文内で動詞と動詞に係っている節を抜き出し，係っている節を適切なフレームに収める．適切なフレームは，助詞によって判断する．例えば「は」が助詞として含まれている場合は主語と判断し subject のフレームへ，「を」が助詞として含まれている場合は目的語と判断し object のフレームに収める．フォーマットの条件を満たしたとき以下のように論理式に変換される

質問文⇒Verb(?subject,?object)

このフォーマットは，動詞や他のフレームによって，入るフォーマットが異なりそれぞれにルールの整備が必要である．

(3) (2)で作成された論理式が条件を満たす時，SPARQL クエリ言語に変換する．

(1) ～(3)の処理を用いた，概念数抽出の例を図 4，概念抽出の例を図 5 に示す．

図 4 では「ある」という動詞に「種類は」と「何種類」が係っている．この時「種類は」には，助詞「は」が含まれているので，subject のフレームに収め，「何種類」には助詞が含まれていないので etc…のフレームに収める．また，それぞれの単語の名称，論理式を次のように SPARQL 言語に変換する．

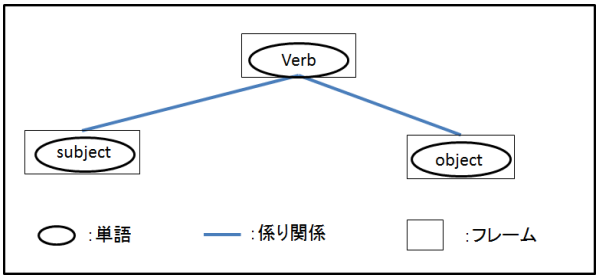

図 3. 基本的な文体系

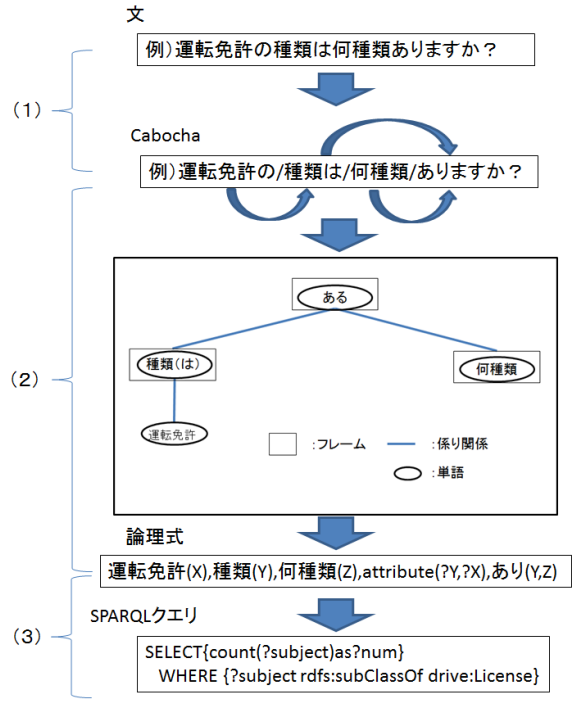

図 4. 概念数抽出までのクエリ変換の例

種類（Y）⇒?subject

運転免許（X）⇒drive:License

Attribute(?Y,?X)⇒subClassOf

あり（?Y,?Z）⇒count()as?num

これら組み合わせる事で SPARQL クエリに変換する．図 5 では「ある」という動詞に「道路標示が」が係っている．この時「道路標示が」には，助詞「が」が含まれているので，subject のフレームに収める．また，それぞれの単語の名称，論理式を以下のように SPARQL 言語に変換する．

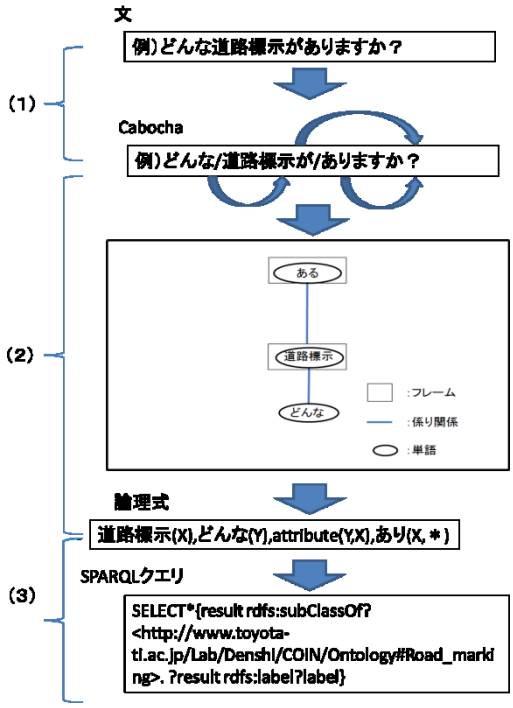

図 5. 概念抽出までのクエリ変換の例

どんな(Y)⇒*

あり (X, ＊) ⇒?result

Attribute(Y,X)⇒subClassOf(Y,X)

道路標示(X)⇒

<http://www.toyota-ti.ac.jp/Lab/Denshi/COIN/Ontology#Road_marking>.

これらを組み合わせる事で SPARQL クエリに変換する.

4. 結果

知識ベースに関しては, 全 162 の単語を収集し, 概念関係を持たせて構築した. 内訳は class が 137 個, Object property が 16 個, Data property が 9 個である . 質問文を SPARQL クエリに変換する事に関しては知識ベース内の全単語のうち次の①②が何単語正しく変換出来たかを評価した.

①X の種類は何種類ありますか？

②どんな X がありますか？

上の形式で X を知識ベース内の全クラスに変えて評価した結果, 137 クラス中 50 のクラスで①, ②に関して共に正解を解答できた.

5. おわりに

137 クラス中 50 クラスの解答という結果は, 最下位クラスの単語や明らかに交通ルールと関係の無い単語を論理式に変換する際に登録しなかった事が原因と考えられる. 今回, 自然言語から成る質問文を知識ベースに問い合わせる際, コンピュータが解釈可能な言語に変換するという問題を, 構文解析器 Cabocha 用い, 論理式に変換する事で解消する事を提案した. しかし, 多様な質問文を解くには, 文章を論理式に変換する為に, 大量のルールの整備が必要であるとわかった. ルールの整備としては「何を問われているのか」のバラエティの充実, 例えば, 「速度は何 km？」「速さはいくつ？」「ある時のスピードは？」が同じ事を聞いていると判断できるように自動車学校の教本や交通冊子内の文章を基にルール作成が必要と考えられ, 知識ベース内の概念数や概念を問う以外にも, 重量や人数の計算, 交通以外の一般常識を問う形にも適応できるようにする必要がある.

参考文献

[1]Tom Kwiatkowski et al., Scaling Semantic Parsers with On-the-fly Ontology Matching. ACL 2013

[2]Cabocha/南瓜:Yet Another Japanese Depency Structure Analyzer (http://code.google.com/p/cabocha/)

[3]Protégé(protégé.stanford.edu)

[4] SPARQL Query Language for RDF (www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/)

[5]警察庁交通局(監修) : 人にやさしい安全運転, 全日本交通安全協会(2011)

[6] 警察庁交通局(監修) : 交通の教則[運転者用], 全日本交通安全協会(2011)

[7]トヨタ名古屋教育センター, 運転教本