

シソーラスを利用したオントロジ統合に基づく知的空間情報検索

齊藤 亮† 田中 文基† 金井 理† 岸浪 建史†

北海道大学大学院工学研究科†

1. はじめに

近年、Web リソースの増加とともに地理的な情報を含む Web リソースが増加し、それらの Web リソースから地理的な情報をメタデータとして取り出し、空間情報検索に利用しようという研究[1]がなされている。しかし、一般的にそれらの研究で用いられるメタデータは、各々異なるオントロジを持つ可能性が高いため、これらの中に潜在的に存在する意味的な関係を活かし、異なるオントロジをもつメタデータを全て対象とした空間情報検索は困難である。

そこで本研究では図 1 のように、WordNet[2]や TGN[3]等のシソーラスと呼ばれる辞書に記述された空間情報に関する用語間の同義、上位/下位、全体/部分などの意味的な関係を SemanticWeb の概念に基づき RDF Schema (RDFS) [4], DAML+OIL[5]で記述することで、異なるオントロジ間にある意味の統合を行い、RDF(S)の検索システム Sesame[6]と上述した意味的な関係を用いて知的空間情報検索を行うことを目的とする。

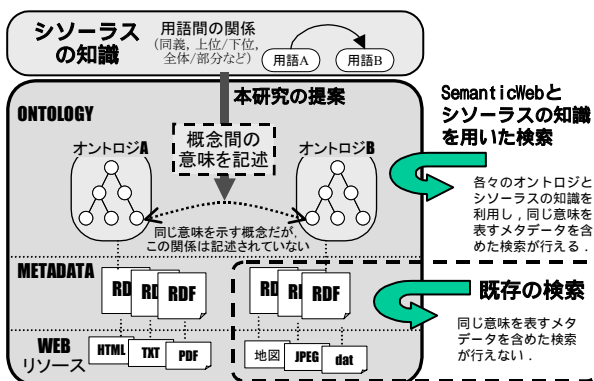


図 1 - 既存の検索手法と本研究での検索手法

2 節では空間情報検索に SemanticWeb の技術を利用する利点についてまとめ、3 節では空間情報に関するシソーラスを利用したオントロジ統合の手法について述べる。4 節では本研究で用いる検索システムの実装方法について述べ、5 節ではその検索システムとシソーラスの知識を利用した知的空間情報検索の例を示す。

2. 空間情報検索への SemanticWeb 技術の適用

W3C の提案する次世代 Web 技術 SemanticWeb[7]では、Web リソースに関するメタデータの記述・交換の枠組み（メカニズム）として RDF[4]を、オントロジやオントロジ間の関係を記述する言語（オントロジ記述言語）として RDF Schema (RDFS)[4], DAML+OIL[5], OWL[8]を提供している。これらのメカニズムや言語を用いることで、あらゆる Web リソースがメタデータを利用し検索可能になり、またメタ

Intelligent spatial information retrieval based on the ontology integration using the thesauruses

† Makoto SAITO, Fumiki TANAKA, Satoshi KANAI, Takeshi KISHINAMI, Graduate School, Hokkaido University, Kita-13 Nishi-8, Kita-ku, Sapporo, Hokkaido, 060-8628, JAPAN

データが表す意味を考慮した高度な検索も可能になる。

この SemanticWeb の技術を空間情報検索へ適用することで、既存の空間情報検索で用いられた技術と比較し以下の利点が期待される。

- 【利点 1】**曖昧性のないメタデータを用いるため、検索結果に不要な情報が含まれない。
- 【利点 2】**検索対象に地理的な情報を含む Web リソースを含めた空間情報検索システムが容易に実装できる。
- 【利点 3】**既存の GIS ではあまり行われていない、オントロジに記述された意味を考慮した検索が可能。
- 【利点 4】**検索システムのオントロジ間に潜在的に存在する意味的な関係を計算機可読な形式で記述することで、複数の検索システムを連携させた検索が可能。

しかしながら SemanticWeb を利用する上で現れる課題には、以下のようなものが挙げられる。

- 【課題 1】**Web リソースに対するメタデータの作成が必須。
- 【課題 2】**オントロジの構築手法。
- 【課題 3】**異なるオントロジの統合手法。

課題 1 に対する研究としては、例えば簡単に Web ページのメタデータの記述が行えるツールの研究[9]がなされており、課題 2 に関する研究は、例えば UML クラス図からオントロジを生成する研究[10]がなされている。しかし空間情報検索においては、異なるメタデータのオントロジ間に存在する意味的な関係を考慮した検索を行う必要がある。そこで本研究では課題 3 を対象とし、シソーラスのような既に体系化された知識に基づき異なるオントロジを統合する手法を提案し、その手法により統合したオントロジを利用し知的な空間情報検索を行う。

3. 空間情報に関するシソーラスを利用したオントロジ統合

3.1 オントロジ間の意味的な関係の記述

空間情報の表現を目的にしたオントロジには、例えば ISO 地理情報規格の UML 図から生成したオントロジ[10]や、様々な地物に関するオントロジを DAML+OIL で表現した研究[11]や、地名カタログサーバの情報を DAML+OIL で表現したオントロジ[12]などがある。これらのオントロジを統合するために、オントロジ間にある意味的な関係を記述するには以下のような知識が要求される。

【要求 1】メタデータの項目（プロパティ）を統合するための体系化された知識。

【要求 2】空間情報を表すメタデータ項目の値となる地名を統合するための体系化された知識。

本研究では、これらの要求を満たすために必要な体系化された知識として、既存のシソーラスを利用することを提案する。

3.2 シソーラスとそのオントロジ統合への適用法

シソーラスは、特定の分野の用語を体系化した辞書であり、辞書内の用語は同義、反意、下位/上位、全体/部分などといった意味により分類・配列がなされている。

空間情報を表すメタデータの項目は、その意味を一般的な名詞の意味で表現できることが多い。従って本研究では、一般的な英語語彙を対象にした WordNet の知識を利用する

ことで要求1を満たす。

また空間情報を表すメタデータの項目は、値に様々な地名を取り得るが、これら地名間にある異名、旧名全体/部分関係などの関係を、地名を対象にしたシソーラス TGN の知識を利用し地名を統合することで、要求2を満たす。

3.3 WordNet[2]

WordNet は、英語語彙データベースとも呼ばれ米 Princeton 大学で作成された英語語彙シソーラスである。WordNet では、同義語の集合を同義語セットまたは synset と呼ばれるグループにまとめることで、用語の概念の定義および分類を行っている。また、全ての開発者がオープンに利用できる。WordNet を利用しある用語に関し取得可能な情報は名詞の場合、表1左列のようなものがある。

本研究では WordNet の version 1.7.1 をプロパティ間にある意味の関係の記述に用いる。

3.4 WordNet によるプロパティ間の意味の記述

WordNet の名詞の用語間にある知識を利用し、異なるオントロジのプロパティ間の意味関係の記述を行うために、本研究では表1のような RDFS, DAML+OIL を用いたプロパティを表す用語間意味関係記述規則を提案する。

表1 WordNet によるプロパティ間の意味の記述規則

用語間の意味の関係	RDFS, DAML+OIL の記述
同義語 Synonym	同等関係を示す daml:samePropertyAs
上位語 Hypernym	汎化/特殊化関係を示す rdfs:subPropertyOf
下位語 Hyponym	
反意語 Antonym	逆向き関係を示す daml:inverseOf
全体 Holonym	(記述しない)
部分 Meronym	(記述しない)
同族語 Coordinate terms	(記述しない)

2つの異なるオントロジ間のプロパティに WordNet における同義語関係が存在する場合、それらのプロパティ間に同等関係を示す daml:samePropertyAs を記述する。同様に、プロパティ間に上位/下位語、反意語関係が存在する場合は、それぞれ rdfs:subPropertyOf, daml:inverseOf を記述する。

図2は、WordNet の名詞「location」に関する知識（「location」は「placement」と同義語であり、「activity」は「location」の上位語である）を利用し、表1の規則に従ってプロパティ「location」と「placement」、「activity」を意味関係でオントロジ統合を行った例である。

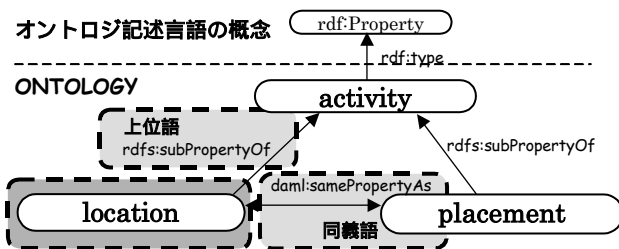


図2- WordNetの知識を利用したプロパティ「location」「placement」「activity」の統合例

3.5 TGN(Getty Thesaurus of Geographic Names)[3]

TGN は、芸術や建築に関する語彙のツールを作成・配布する Getty プログラムの一つであり、約 1,000,000 の地名を全体/部分関係を用い階層的にデータ化している。TGN (July, 2000 発行のもの) により取得可能な情報は表3左列のような異名や旧名、地理的な全体/部分関係、PlaceType などがある。

PlaceType とは、その地名のタイプにあたるものであり、一つの地名には複数の PlaceType が指定されても良い。PlaceType の値には例えば表2のようなものがあるが、それらの PlaceType の値の間には上位/下位といった階層関係は記述されていない。

表2 PlaceType の値の例

物理的区分	行政・機能的区分
<ul style="list-style-type: none"> Continent Island Sea 	<ul style="list-style-type: none"> Depression Valley Canyon Nation Kingdom Capital port City Market center

3.6 TGN による地名を表すクラス間の意味の記述

TGN の異名、旧名、全体/部分関係の知識を利用し、異なるオントロジの地名を表すクラス（地名クラス）間に潜在的に存在する意味関係の記述を行うために、本研究では表3に示されるような地名クラス間に RDFS, DAML+OIL を用いた意味関係の記述規則を提案する。

表3 TGN による地名クラス間の意味関係の記述規則

地名に関し得られる情報	RDFS, DAML+OIL による記述
異名, 旧名	daml:sameClassAs
地理的な全体部分関係	地理的な全体部分関係のそれぞれに、daml:disjointUnionOf と daml:collection を用い表現する
PlaceType	クラス (rdfs:Class) として各値を定義し、対象とする地名をこのクラスの rdfs:subClassOf とする。
座標	(記述しない)
地名が記された出所	(記述しない)
その他 (メタデータの作者など)	(記述しない)

TGN に含まれる地理的な全体部分関係は、DAML+OIL の「互いに素である列挙された参照クラスの和集合をとるプロパティ daml:disjointUnionOf」と集合の概念 daml:collection を組み合わせて表現した全体/部分関係にマッピングされる。例えば表3の規則に従い地名クラス「Hokkaido」、「Ezo」、「Sapporo」、「Hakodate」間の意味関係でオントロジ統合を行った例は図3のようになる。

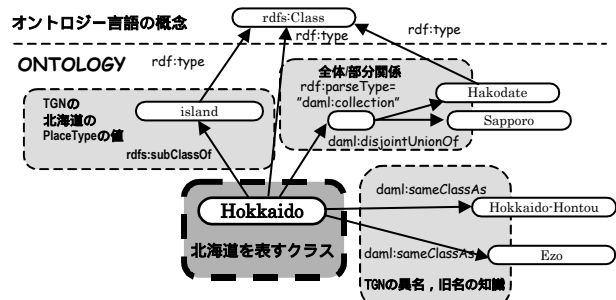


図3-TGNの知識を利用したクラス「Hokkaido」のオントロジ統合

4.空間情報検索システムの実装

4.1 検索のためのクエリ言語とツール

DAML+OIL で記述したオントロジに対するクエリ言語には DQL (DAML Query Language) [13] があり DQL を実装した検索ツールには DQL tool kit [13] がある。これらは、最近 (2002.8) 公開されたばかりであるため、クエリ言語、検索ツールともにまだ機能が乏しい。また RDFS で記述されたオントロジに対するクエリ言語には RQL (RDF Query Language) [14] があり、RQL を実装した検索ツールには Sesame [6] がある。一方、RDF を検索することができる検索ツールには RDFDB [15], Squish [16], Jena RDQL [17] などがあがるが、これらは直接三つ組みのステートメントとして表現されるリソースしか検索できない。それに対し DQL や RQL を用いた検索ツールでは、意味を考慮した推論 (推移律など) による検索が可能である。例えば「概念 A の特殊化概念が概念 B である」と「概念 B の特殊化概念が概念 C である」という記述から、「概念 A の特殊化概念」の検索結果として「概念 B」と「概念 C」が検索できる。

そこで本研究では、機能の豊富さの理由により検索ツールには Sesame を用いることとした。

4.2 検索ツール Sesame の機能

Sesame は、European IST project On-To-Knowledge で開発された RDF(S) の検索システムであり、その検索エンジンには RDF(S) のクエリ言語 RQL を実装していたため、レポジトリ内に格納されたオントロジやメタデータに対し上述のような推論を利用した複雑な検索ができる。また Sesame では部分的に DAML+OIL をサポートしている。さらに Sesame には Web ブラウザを用いて簡単な検索が行える GUI も備えている。

4.3 全体/部分関係の検索

3.5 節で述べたように TGN にある全体部分関係は、DAML+OIL のプロパティ `daml:disjointUnionOf` と集合の概念 `daml:collection` を用いて記述した。しかしながら、このように記述された集合に対する検索は、現在の RQL では行うことが出来ない。

そこで本研究では、Sesame に `daml:collection` をサポートした RDF パーサ ARP (Another RDF Parser) [18] を組み合わせることで、図 4 左のような全体部分関係を、図 4 右のように展開した全体部分関係を格納し、検索可能にした。

格納された全体部分関係は、いくつかのステートメントの集合として表現され、複雑な構造となっているため、全体部分関係の検索は容易に行えない。そこで本研究では、容易に検索するため Web ブラウザにて RQL 式の生成を行うツール (図 5) を JavaScript にて実装し、これを Sesame のフロントエンドとともに利用し検索を行う。なお、現在のこのツールでは、直接全体部分関係を持つ地名のみしか検索できない。

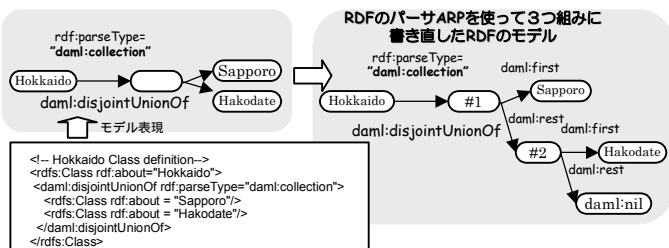


図 4 全体部分関係のRDFパーサARPによる展開

5.提案手法を用いた知的空間情報検索の例

シソーラスの知識を利用した 2 つの基本的な検索の例を示す。ただし、この節で示すメタデータやオントロジで利用される地名は、例えば図 6 や図 7 のように、インスタ

スおよびクラスの両方として既に作成されているものとし、それらクラス-インスタンス間の関係を `rdf:type` で定義されているものと仮定した。

5.1 クラス、プロパティの同等性を利用した検索

図 6、図 7 は北海道庁の Web ページのメタデータとオントロジであり、図 8、図 9 は五稜郭の JPEG 写真に関するメタデータおよびオントロジである。

図 7 のオントロジで定義されたプロパティ「location」と図 9 のオントロジで定義されたプロパティ「placement」に対し、WordNet の同義語の知識を利用しプロパティの同等関係を図 2 のように記述した。また図 7 のオントロジで定義された地名クラス「Hokkaido」と図 9 のオントロジで定

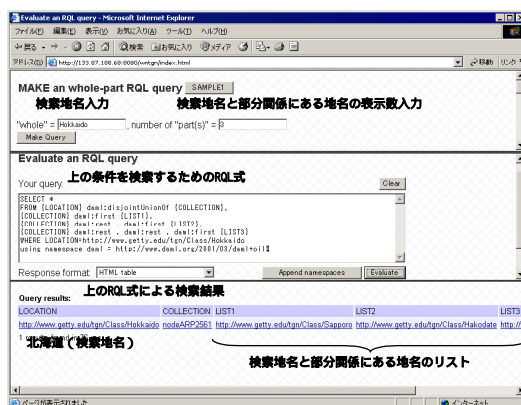


図 5 全体部分関係検索用RQL式生成ツール

```

<!-- DouCHO -->
<rdf:Description rdf:about="http://www.pref.hokkaido.jp/">
  <dc:title>Hokkaido Prefecture</dc:title>
  <wn:location>
    <rdf:Description rdf:about="http://.Instance/Hokkaido">
      <rdf:type resource="Hokkaido"/>
    </rdf:Description>
  </wn:location>
</rdf:Description>
    
```

図6-道庁Webページメタデータ

```

<!-- Goryokaku jpeg -->
<rdf:Description rdf:about="http://.....jp/~juta/casting/syasinsyu/goryokaku.jpg">
  <dc:title>ancient Goryokaku fortress</dc:title>
  <wn:placement>
    <rdf:Description rdf:about="http://www.getty.edu/tgn/Instance/Ezo">
      <rdf:type resource="Ezo"/>
    </rdf:Description>
  </wn:placement>
</rdf:Description>
    
```

図8-五稜郭JPEG写真メタデータ

```

<rdf:Property about="location"/>
<rdf:Class about="Hokkaido"/>
    
```

図7-北海道庁Webページオントロジ

図9-五稜郭JPEG写真オントロジ

```

SELECT URI, @SHOZAI, TGNInstance
FROM (URI) @SHOZAI (TGNInstance), (TGNInstance) rdf:type (TGNClass)
WHERE @SHOZAI >= location and @SHOZAI <= location and
TGNClass = Hokkaido
    
```

図10-概念の同等関係を利用した検索のRQL式

Query results:

URI	@SHOZAI	TGNInstance
http://www.pref.hokkaido.jp/	location	http://www.getty.edu/tgn/Instance/Hokkaido
http://www....insyu/goryokaku.jpg	placement	http://www.getty.edu/tgn/Instance/Ezo

図11-概念の同等関係を利用した検索結果

義された地名クラス「Ezo」に対し、TGN の旧名に関する知識を利用し、地名クラス間の同等関係を daml:sameClassAs で図 3 のように記述した。

これらの RDF(S) や DAML+OIL で記述したメタデータやオントロジ、オントロジにあるプロパティ間やクラス間の関係を Sesame へ格納し、図 10 の RQL 式にて検索を行った。この RQL 式では、「プロパティ location」およびそれと同等のプロパティの示す値が、「Hokkaido」もしくはそれと等価な地名となっている Web リソースの URI が検索される。検索結果は図 11 のようになり、プロパティ「placement」と地名クラス「Ezo」で記述された図 8 メタデータの示す Web リソースの URI が同時に検索されている。従ってシソーラスの知識によってオントロジ間に記述したプロパティやクラスの同等関係により、記述の異なる意味的に等しいメタデータの検索が行えることが分かる。

5.2 全体/部分関係をもつ地名の検索

TGN の地名に関する全体/部分関係の知識を利用することで、オントロジとして定義された地名クラス「Hokkaido」、「Sapporo」、「Hakodate」間の関係(図 3)を、DAML+OIL で記述し、Sesame へ格納した。

図 5 に示す全体部分関係の RQL 式生成ツールを用いて、クラス「Hokkaido」と部分関係にある地名を表すクラス「Sapporo」や「Hakodate」の検索を行った。図 12 はクラス「Hokkaido」と部分関係となる地名クラスを検索する RQL 式であり、その検索結果は図 13 のようになり、部分関係となる地名クラスが正しく検索されたことが分かる。

6. おわりに

本研究では、WordNet や TGN 等のシソーラスに記述された空間情報に関する用語間の同義、上位/下位、全体/部分などの意味的な関係を、SemanticWeb の概念に基づき記述することで、異なるオントロジ間にある意味の統合を行う手法を提案し、RDF(S)の検索システム Sesame とシソーラスの知識とを用いて手法の実装を行い、提案する知的空間情報検索手法が有効であると確認した。

今後さらに、PlaceType の意味を考慮した検索機能を組み合わせることで、より柔軟な空間情報検索を実現する予定である。

```

SELECT *
FROM {LOCATION} daml:disjointUnionOf {COLLECTION},
      {COLLECTION} daml:first {LIST1},
      {COLLECTION} daml:rest . daml:first {LIST2},
      {COLLECTION} daml:rest . daml:rest . daml:first {LIST3}
WHERE LOCATION= Hokkaido

```

図12-生成した全体部分関係検索用RQL式

参考文献

- [1] 河合英紀ら, Support Vector Machine を用いた地域情報ページの自動分類, FIT2002 講演論文集, 2002.9
- [2] WordNet, <http://www.cogsci.princeton.edu/~wn/>
- [3] TGN, <http://www.getty.edu/research/tools/vocabulary/tgn>
- [4] RDF, <http://www.w3.org/RDF/>
- [5] DAML+OIL, <http://www.daml.org/>
- [6] Sesame, <http://sesame.aidadministrator.nl/>
- [7] IPSJ, 情報処理 vol.43, No.7, 特集セマンティック Web, pp707-750, 2002
- [8] OWL, <http://www.w3.org/2001/sw/WebOnt/>
- [9] Annotea, <http://www.w3.org/2001/Annotea/>
- [10] 斉藤ら, ISO 19100 地理空間データの SemanticWeb での利用に関する研究, FIT2002 講演論文集, 2002.9
- [11] DAML GeoFile, <http://www.daml.org/2001/02/geofile/>
- [12] GeoNames in DAML, <http://www.daml.org/2002/04/geonames/>
- [13] DQL, <http://ksl.stanford.edu/projects/dql/>
- [14] RQL, <http://139.91.183.30:9090/RDF/RQL/>
- [15] RDFDB, <http://www.guha.com/rdfdb/>
- [16] RDF Squish query language, <http://ilrt.org/discovery/2001/02/squish/>
- [17] Jena, <http://www-uk.hpl.hp.com/people/bwm/rdf/jena/>
- [18] ARP, <http://www-uk.hpl.hp.com/people/jjc/arp/>

Query results:		
LOCATION	LIST1	LIST2
http://www.../Class/Hokkaido	http://.../Class/Sapporo	http://w.../Class/Hakodate

図13-生成した全体部分関係検索による検索結果