

オントロジー工学的観点に基づく Semantic Web アプリケーションの傾向分析

An Analysis of Semantic Web Applications based on Ontological Engineering

古崎 晃司
Kouji Kozaki

林 雄介
Yusuke Hayashi

笹島 宗彦
Munehiko Sasajima

溝口 理一郎
Richiro Mizoguchi

*¹ 大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Ten years have passed since the concept of the semantic web was proposed by Tim Berners-Lee. For these years, basic technologies for them such as RDF(S) and OWL were published. As the result, many systems using semantic technologies have been developed. Some of them are not prototype systems for researches but real systems for practical use. The authors analyzed semantic web applications published in the semantic web conferences (ISWC, ESWC, ASWC) and classified them based on ontological engineering. This article discusses a trend and the future view of them using the results.

1. はじめに

Tim Berners-Lee によって Semantic Web の基本的な概念が提唱されてより約 10 年を経て、RDF(S)や OWL といった基盤技術の整備が進むと共に、多くのアプリケーションが開発されるようになった。この間 Semantic Web に関連する研究プロジェクトも数多く実施され、研究レベルのプロトタイプのみならず、実用レベルの大規模システムの開発も盛んに行われている。企業での Semantic Web 技術の利用も本格化されつつあり、既に約 300 社の企業が Semantic Web 技術を提供しているというレポートもある[Mills 08]。しかし、“Semantic Web のキラーアプリケーションは何か？”という問いは今もなお聞かれ[Alani 05]、より有用性の高い Semantic Web アプリケーションの開発を目指し、多種多様な領域を対象とした研究・開発が盛んに行われている。

本論文では、このような Semantic Web 技術を利用した応用システムに関する研究動向を、オントロジー工学的な観点から分析・考察し、現状の Semantic Web 応用研究の課題と将来の展望について議論する。具体的には、Semantic Web 分野で中心的な国際会議である ISWC, ESWC および ASWC で発表された論文を対象として、用いられているオントロジーの種類、システムの利用目的、対象領域という 3 つの観点から分析を行う。

2. 分析対象

本論文では、Semantic Web 分野で最も中心的な国際会議である ISWC, ESWC および ASWC において、これまでに発表された論文 (Industrial セッションなども含む本会議のみ) を対象とした。これらの中から、筆者らの判断で Semantic Web の応用システムに関するものと思われる論文をすべて選び出し、合計 190 編を今回の分析の対象とした。なおアプリケーションに関する論文であっても、オントロジー構築ツールなど直接的な応用システムとは言えないものや、理論的な枠組みのみを述べて実際のシステムが開発されていないと思われるものについては基本的に対象外とした。以下に各会議の開催情報と、それぞれで分析対象となった論文数を示す。

- **International Semantic Web Conference (ISWC)**
 - ISWC2002, Jun. 9-12, 2002, Sardinia, Italy……………9
 - ISWC2003, Oct.20-23,2003, Sanibel Island, FL, USA…19
 - ISWC2004, Nov. 7-11, 2004, Hiroshima, Japan……………18

- ISWC2005, Nov. 6-10, 2005, Galway, Ireland……………25
- ISWC2006, Nov.5-9, 2006, Athens, GA, USA……………26
- ISWC2007,Nov.11- 15, 2007, Busan, Korea……………18
- **European Semantic Web Conference (ESWC)**
 - ESWC2005, May29-Jun.1, 2005, Heraklion, Greece…24
 - ESWC2006, Jun.11-14, 2006, Budva, Montenegro……11
 - ESWC2007, Jun. 03 - 07, 2007, Innsbruck, Austria……17
- **Asian Semantic Web Conference (ASWC)¹**
 - ASWC2006, Sep.3- 7, 2006, Beijing, China……………23

3. Semantic Web アプリケーションの分析手法

Semantic Web の基本的なコンセプトは、Web 上のリソースにオントロジーで定義した語彙を用いた計算機可読のメタデータを付与して、意味処理を可能とする点にある。よって、オントロジーは Semantic Web アプリケーションの基盤技術の 1 つであり、“どのようなオントロジーを用いているか”が、分析の重要な観点となる。どのようなオントロジーを利用するかは、そのアプリケーションの利用目的に応じて必要とされる意味処理に依存する。そこで本研究では、Semantic Web アプリケーションの「**利用目的(タイプ)**」と用いられている「**オントロジーの種類**」から分析を行った。ここで分析に用いた利用目的とオントロジーの種類の分類指標には、国内のオントロジー利用研究の傾向分析[古崎 07]で考察した分類を、Semantic Web アプリケーション向けに再検討したものを利用した。以下に、それぞれの概要を述べる。

3.1 Semantic Web アプリケーションの利用タイプ

Semantic Web アプリケーションにおける意味処理は、オントロジーに基づいて行われる。よって、アプリケーションの利用目的を、オントロジーがどのような目的で利用されるかを中心に分類した。本論文では従来の分類[古崎 07]に、「検索」、「分析」、「抽出」の 3 つのタイプを追加した。また今回の分析では、各アプリケーションをいずれか 1 つのタイプに分類するのではなく、複数のタイプに分類することを許容した。これにより、そのアプリケーションの利用目的をより正確に表すことが出来ると考えた。

(1) 共通語彙としての利用

知識の相互性を高めるために、知識を記述する際に用いる共通語彙としてオントロジーを利用する。最も基本的なオントロジーの利用法であるため、以下で述べる(2)~(9)の利用タイプ

連絡先: 古崎晃司, 大阪大学産業科学研究所, 〒567-0047 大阪府茨木市美穂が丘 8-1, Tel:06-6879-8416, Fax:06-6879-2123, kozaki@ei.sanklen.osaka-u.ac.jp

¹ 第 2 回にあたる ASWC2007 は ISWC2007 と共同開催された。

においても大前提となる利用法であるが、特に語彙の統一を中心の役割としているシステムのみをこの分類に入れる。

(2) 検索

適切なメタデータを付与することで、単なるキーワードマッチによる全文検索などでは実現できない、意味処理を含む検索が可能となる。どの程度まで詳細な意味を扱えるかは、付与するメタデータと基となるオントロジーの詳細度などに依る。Semantic Web アプリケーションの最も典型的な利用法であるが、特に検索がシステムの中心と思われる場合にこの分類に入れる。

(3) インデックスとしての利用

適切な情報へアクセスのためのインデックスとしてオントロジーを利用する。この利用法は、オントロジーが提供する語彙体系をインデックス(見出し)として利用したもので、その語彙を用いて情報にアクセスする。体系の中における語彙の位置づけを明示的に利用する点が単なる共通語彙としての利用や検索とは異なる。意味処理を利用した情報アクセスのためのナビゲーションなどもこの分類に入れる。

(4) データスキーマとしての利用

データベースなどに格納するデータの構造や値を規定する標準化されたデータスキーマとしてオントロジーを利用する。オントロジーが提供する概念階層がデータの値となるインスタンスを規定し、各概念定義がデータの構造を規定する。典型的な利用例としては、データベースにおけるスキーマや、ドキュメントに付与する書誌情報などといったメタデータの内容を規定するメタデータスキーマとしての利用などがある。

(5) 知識共有の媒体としての利用

オントロジーおよびオントロジーに基づくインスタンスモデルを用いて、システム間、人間とシステム、あるいは人間同士の間で知識の相互運用性を高め知識共有を実現する。オントロジーの利用は一般に知識共有の実現を目指したものと言えるが、ここでは複数システム間での知識の変換やマッピング、エージェント間の通信、コミュニケーション支援など、知識の相互運用性を中心的な目的とした利用を、このタイプに分類する。技術的には図 1(a)に示すように、複数の知識を同一のオントロジー(参照オントロジー)と対応付けることで変換やマッピングを行う手法や、図 1(b)のように複数のオントロジーを直接比較する手法などがある。

(6) 分析

メタデータを単なる検索に利用するだけではなく、オントロジーに基づく推論・意味処理の技術などを用いて、メタデータが指す内容を分析することができる。最も典型的な分析手法には、オントロジーの推論機構を用いた自動分類がある。このように検索結果を何らかの手法で分析している場合、この分類に入れる。

(7) 抽出

メタデータなどを分析するだけでなく、分析結果に基づいて利用者にとって意味のある情報を抽出することを目的とするシステムをこの分類に入れる。利用者の嗜好に合わせた情報を提示する情報推薦システムも、抽出の一種に位置づける。

検索を例に取ると、検索結果をそのまま提示する場合は(2)、何らかの分析結果を併せて提示する場合は(6)、結果を絞り込んで意味のある情報を抽出している場合は(7)に分類する。

(8) 知識モデルの規約としての利用

オントロジーが提供する概念定義に基づいて構築されたインスタンスモデルは、対象世界のオブジェクトを表した知識モデル

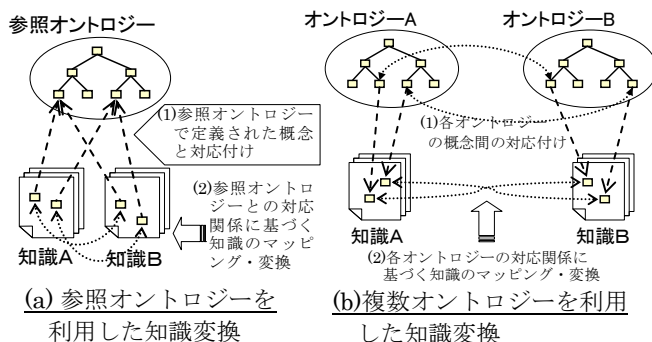


図 1. 知識共有の媒体としての利用

として用いることができる。すなわちオントロジーは、その知識モデルに対する規約を与えるメタモデルとして利用できる(図 2)。これは(4)で述べたデータとデータスキーマの関係に相当するが、知識モデルには単なるデータに比べて柔軟な知識の意味記述が求められる。



図 2. 知識モデルの規約としての利用

オントロジー工学的立場からは、この知識モデルの規約として利用できることが、オントロジーの最も本質的な役割と位置づけられる。対象世界を適切に捉えたヘビーウェイト・オントロジーを構築・利用することで、より深い知識を計算機で扱えるようになるが、広く浅い知識を対象とするときなど、情報処理の効率を重視してライトウェイト・オントロジー¹が好まれる場合もある。

(9) 知識の体系化への利用

オントロジーは知識の体系化・組織化の核となる概念構造を提供する。それらを用いることで、対象世界の知識を構成している概念をオントロジーとして体系化し、その概念定義が与える規約に基づいて知識管理・運用を実現することができる。これは(1)~(8)で述べてきたオントロジーの利用目的を総合的に用いたものと位置づけることができ、知識管理システムやコンテンツ管理システムといった統合的なシステムが、その典型的な応用例となり得る。

3.2 オントロジーの種類

本論文ではオントロジーの種類を、対象とするドメインや記述に用いる言語に依らない共通の分類の観点として、「用いられているオントロジーの意味的構成要素」によって以下の 5 種類に分類する。

(A) 簡単なスキーマ

本来はオントロジーとは呼ばれないものであるが、Semantic Web アプリケーションにおいては、RSS や FOAF のように統一したデータ記述のために簡単なスキーマが用いられる場合がある。

(B) 概念間の is-a 関係に基づく概念階層

対象世界に存在する概念を切り出した結果としての「概念」の集合と、それらの概念間の is-a (上位-下位) 関係 (RDF(S) や OWL では rdfs:subClassOf で表される) を同定することによる階層化が、オントロジーにとって最も本質的な要素となる。Web 検

¹ オントロジー工学的立場からは対象世界に対する考察の深さの違いで、ヘビーウェイト/ライトウェイト・オントロジーを区別するが、Semantic Web 分野では公理などのクラス定義の記述量でヘビーウェイト/ライトウェイトが区別されることが多い。

素サイトのトピック階層など、is-a 階層のみから成るオントロジーをライトウェイト・オントロジーと呼ぶ場合もある。

(C) その他 (is-a 以外) の関係を含む

is-a 関係に基づく概念の階層化に加え、それ以外の関係を用いることで各概念の定義をより明確に表すことができる。part-of(全体一部分)関係や、attribute-of(属性)関係が最もよく使われる関係である。RDF(S)や OWL のプロパティは、このような関係を記述するのに用いられる。

(D) 意味制約の公理的記述を含む

関係を用いた各概念の意味定義の記述に加え、意味制約の公理的記述を行うことで、インスタンスモデルにおいて適切な規約を与えることができる。ここで記述される意味制約には、推移律や反射律などの関係の性質、インスタンスの排他性に関する制約などが含まれる。OWL を用いた場合は transitiveProperty, inverseOf といった関係の性質に関する制約や disjointWith, oneOf といった意味制約を用いることができる。

(E) その他の強い公理¹を含む

意味制約だけでは十分に表しきれない意味定義に関しては、ルール記述言語などを用いた強い公理とした記述が必要となる。KIF や SWRL などを用いたルール記述を伴うオントロジーがここに分類される。

3.3 分析の手順

- 2章で述べた対象論文 190 編に対して、著者ら 3 人で分担し、
- システムの概要 (1 文程度での説明)
 - 利用タイプ (3.1 節の分類: 複数選択可)
 - 対象としているドメイン
 - オントロジーの種類 (3.2 節の分類: いずれか 1 つ選択)
 - 利用しているオントロジー記述言語
 - オントロジーの概念数およびインスタンス数

を論文から読み取れる範囲で調べた。作業の中間段階において、お互いの作業上で生じた問題点について議論し分類方針の調整を行った。3.1 節で述べた利用タイプの分類は、この議論の過程で調整した結果を示している。

4. 分析結果と考察

4.1 利用タイプの分布

図 3 のグラフは対象論文の全システムが、それぞれの利用タイプに分類された数の分布を示す。グラフより、全体としてはすべての利用タイプに対して、ほぼ同程度の割合で分類されていることが見て取れる。ただし、検索をはじめとしたデータ処理が中心の(1)~(7)と、オントロジーに基づいた知識処理を中心とした(8), (9)の割合を比較すると、現状の Semantic Web アプリケーションの多くは構造化されたデータ処理が中心であることが伺える。

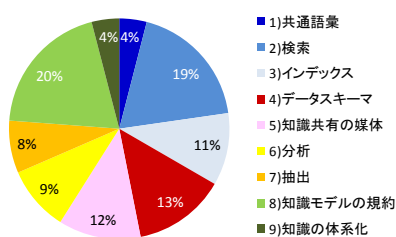


図 3. 利用タイプの分布

¹ 公理やルールという言葉が表す内容には、意味定義に用いる制約や、推論に用いる規則など多くの意味が含まれている。ここでは厳密な議論を避け、前述の 1~3) で表される内容以外を「強い公理」と呼ぶ。

表 1. 利用タイプとオントロジーの種類

	オントロジーの種類					合計
	簡易スキーマ	概念階層	その他の関係	意味制約	公理あり	
1) 共通語彙	0	4	7	0	0	11
2) 検索	1	2	43	4	1	51
3) インデックス	0	3	23	3	0	29
4) データスキーマ	0	0	32	5	0	37
5) 知識共有の媒体	1	0	31	1	0	33
6) 分析	1	1	21	3	0	26
7) 抽出	1	2	15	3	0	21
8) 知識モデルの規約	0	1	36	9	8	54
9) 知識の体系化	0	2	8	1	0	11
合計	4	15	216	29	9	273

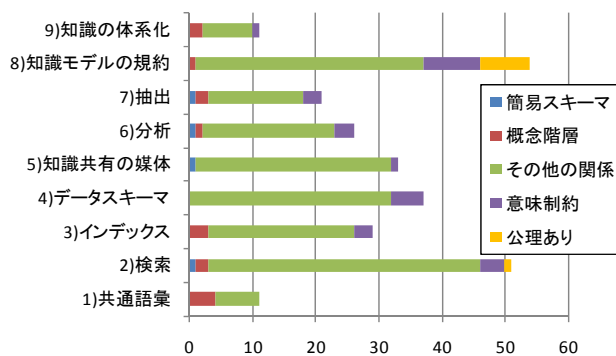


図 4. 利用タイプとオントロジーの種類

4.2 利用タイプとオントロジーの種類の間

表 1 および図 4 に利用タイプ毎に用いられているオントロジーの種類を示す。この表より、多くの Semantic Web アプリケーションは is-a 関係の他に必要な関係を定義したオントロジーを用いており、複雑な公理記述を含むものは少ないということが示唆される。ただし、オントロジーに意味制約の公理的記述が含まるか否かを論文のみで判断するのは難しく、今後更なる分析が必要と思われる²。また、利用タイプが(1)~(9)へとより複雑な意味処理を伴うものになるにつれて、用いるオントロジーもより詳細な記述が必要であることが読み取れる。

使用されているオントロジー記述言語の割合 (図 5) は半数が OWL もしくは OWL を拡張した言語、次いで RDF(S) となっており、OWL の Semantic Web 用のオントロジー記述言語の標準として位置づけが固まりつつあることが分かる。なお OWL の Lite/DL/Full のいずれを使用しているかについては明記されていない論文が多く、残念ながら正確な数が把握できていない。

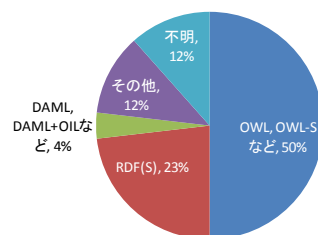


図 5. オントロジー記述言語の使用割合

4.3 会議毎の利用タイプの推移

図 6 に各利用タイプに分類されたシステムの数の会議毎の推移を示す。会議は開催年月順に並べているので、約 5 年間に渡る Semantic Web アプリケーションの開発動向の概要を知ることができる。全体の傾向としては、ISWC2002 と ESWC2006 を除き分析対象となった論文の件数は 20 前後で大差はないが、利用タイプの総数は右肩上がりとなっていることが見て取れる。これは、近年になるにつれて Semantic Web アプリケーションの

² その他の強い公理を含むか否かは、SWRL などルール記述言語の使用を明示しているか否かで判断した。

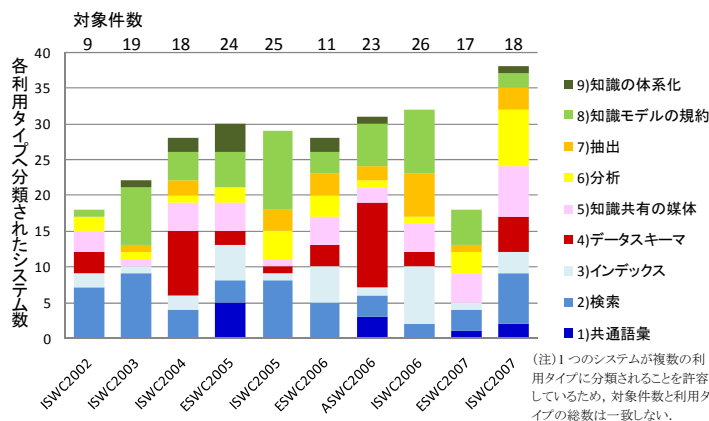


図 6. 会議毎の利用タイプの推移

研究開発が成熟し、より多くの機能を持ったシステムが開発されるようになり、複数の利用タイプに分類されているためと思われる。特に、(1)~(3)の語彙や検索のための基本的な利用タイプの数には大きな変化が見られないが、(4)~(9)で示されるより高度な意味処理の利用が増加している。

しかし、4.1 節で議論したように、(8)、(9)で示されるオントロジーに基づいた知識処理を中心とした利用はさほど増加していない。これは、深い知識処理を有効に利用したアプリケーションの開発の難しさを示唆していると思われる。この点については、データ処理を中心としたアプローチが Semantic Web アプリケーションの主流であるという捉え方と、逆に、深い知識処理の難しさを克服することがキラーアプリケーションを生み出す糸口であるという捉え方ができる。

4.4 対象ドメイン

図 7 に対象ドメイン毎の利用タイプの分布を示す。多くのアプリケーションが開発されているドメインとしては、マルチメディア (画像、動画、音楽など)、サービス (Web サービス、現実世界のサービスの両方を含む)、ソフトウェア開発、ナレッジマネジメント、バイオインフォマティクス、医療などがある。ビジネス分野のドメインは、製品管理、ビジネスプロセス、自動車などより対象を絞ったシステムが幅広く開発されていた。また特定ドメインを対象としない汎用のシステムも多く開発されているが、図 7 では、論文や会議情報などの学術情報一般を対象とするもの、Web ページをはじめとする Web リソースを主な対象とするもの、それ以外の汎用のフレームワークを提供するものに分類した。

利用タイプの分布については、ドメイン毎の大きなばらつきは見られず、Semantic Web 技術が様々な分野で広く利用されていることが分かる。

4.5 オントロジーの概念数およびインスタンス数

アプリケーションの規模に応じて、オントロジーの規模やメタデータとして付与されるインスタンス数は大きくなるのが一般的に予想される。今回の分析で対象とした論文には、オントロジーの概念数やインスタンス数が明記されているものが少なく、十分な分析を行うことができなかった。

論文から読み取れる範囲では、特に近年に開発されているアプリケーションでは、DOLCE¹や WordNet²、DBpedia³など既存の大規模オントロジー (またはソーラス) などを利用したシステムが多く見られる。また、独自のオントロジーを利用している場合も論文では概要が述べられているのみで、その規模や定義

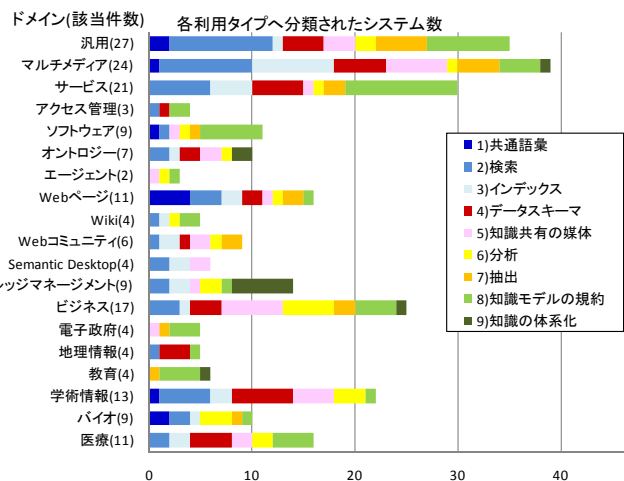


図 7. 対象ドメイン毎の利用タイプの分布

内容の詳細を十分に分析はできていない。構築したオントロジーを Web 上で公開している場合も見られるので、今後より詳細な分析を行う際には、それらをダウンロードして内容を確認するという方法も残されている。

5. まとめ

本論文では、ISWC, ESWC および ASWC で発表された論文を対象として、Semantic Web アプリケーションの研究・開発動向の分析を行った。既に様々な分野で多くのシステムが開発されており、特に近年は、比較的大規模なアプリケーションが実用に近いレベルで開発されていると言っても過言ではない。Semantic Web が提唱された当初のような“全世界の Web”を対象とした意味処理の実現はまだほど遠いが、“ある程度の規模”の Web やローカルシステムでの Semantic Web 技術の利用は既に実現しつつある。

今後の Semantic Web 技術を用いたキラーアプリケーション開発に向けての大きな課題の 1 つは、真の意味での Semantic をどのように実現するかにあると思われる。オントロジー工学の研究者としては、オントロジーに基づいた深い知識処理が、その突破口に成ることを目指して、研究を進めたい。

なお、今回の分析を通して、欧米を中心に多くの Semantic Web 関連のプロジェクトが幅広く展開していることを、改めて実感させられた。同時に、国内の研究開発の遅れに多少の懸念を感じると共に、本論文の分析結果⁴が今後の Semantic Web 発展の一助となれば幸いである。

参考文献

[Davis 08] Mills Davis: Semantic Wave 2008 Report: Industry Roadmap to Web 3.0 & Multibillion Dollar Market Opportunities, www.project10x.com, 2008.
 [Alani 05] Harith Alani, Yannis Kalfoglou, Kieron O'Hara, and Nigel Shadbolt: Towards a Killer App for the Semantic Web, Proc. of 4th International Semantic Web Conference (ISWC 2005), Galway, Ireland, November 6-10, pp. 829-843, 2005.
 [古崎 07] 古崎晃司, 来村徳信, 溝口理一郎: オントロジー工学的立場から見たオントロジー利用研究の分類と考察, 第 21 回人工知能学会全国大会, 1D3-3, 2007.

¹ <http://www.loa-cnr.it/DOLCE.html>

² <http://wordnet.princeton.edu/>

³ <http://dbpedia.org/>

⁴ 分析結果の詳細は、

http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_apps/ に公開を予定している。