

小特集 「国際会議で見つけたオススメ論文」

The 5th International Semantic Web Conference (ISWC2006)

森 純一郎
Junichiro Mori

東京大学大学院情報理工学系研究科, ドイツ人工知能研究所
Graduation School of Information Science and Technology, The University of Tokyo. / German Research Center for Artificial Intelligence (DFKI).
jmori@mi.ci.i.u-tokyo.ac.jp

Keywords: semantic web, rdf, sparql.

1. 紹介論文の背景

本稿では、昨年2006年11月に開催されたセマンティック Web に関する国際会議 (ISWC2006) から SPARQL に関する論文を紹介する。SPARQL (Simple Protocol And RDF Query Language)^{*1} は、RDF のためのクエリ言語である。SPARQL は2004年10月に W3C のワーキングドラフトとして発表され、現在は W3C の勧告候補になっている。RDF のクエリ言語の統一された仕様として SPARQL は、セマンティック Web に関する研究の題材として近年取り上げられてきている。

本稿で紹介する論文「Semantics and Complexity of SPARQL」[Jorge 06] では、SPARQL の意味論について、構成的方法に基づく提案を行い、クエリ言語としての SPARQL の複雑性や最適化に関する考察を行っている。また本稿では、SPARQL の拡張について述べている論文として「Querying the Semantic Web with Preferences」[Wolf 06] の紹介も行う。前者の論文が、意味論や複雑性など SPARQL の基礎的な部分を論じているのに対して、後者は SPARQL の応用として Preference を含むクエリの拡張を論じている。そのため、これら二つの論文は現在の SPARQL に関する基礎から応用にわたる研究動向を概観するのに適切なものである。

なお、これら2編の論文は先の ISWC 2006 において、それぞれ Best research paper award, Best student paper award を受賞しており、セマンティック Web コミュニティにおける SPARQL に対する近年の注目の高まりがうかがえる。

2. Jorge, P., Marcelo, A. and Claudio, G.: Semantics and Complexity of SPARQL

2.1 研究の背景と目的

SPARQL は、RDF グラフのマッチングに基づくクエリ言語である。SPARQL クエリ (以下、本稿ではグラフパターンとする) は、Pattern matching, Solution modifiers, そして Output という主に三つのパートからなる。SPARQL を構成する個々の要素は単純であるが、それらの組合せは SPARQL を複雑な言語にしている。現在の SPARQL の意味論 [spa] においては、SPARQL の構成要素の組合せがもたらす複雑性に対する考察がなされていない。これらは W3C のメーリングリストでも長い議論が行われている。

本論文 [Jorge 06] では、SPARQL の意味論について形式的な考察を行う。特に、構成的な意味論を提案したうえで、複雑性について、SPARQL グラフパターン表現の評価が多項式領域 (PSPACE) 完全であることを示している。また、従来の操作手続き的な意味論との比較を行い、ある条件のもとでは、従来の意味論と提案手法が等価であることを示している。さらに、SPARQL における正規形や最適化についての議論を行う。

これまでの研究においては、SPARQL の一部の構成要素に焦点を当てた部分的な意味論の提案や、より一般的な RDF クエリ言語に関する意味論についての提案がなされてきた。一方、本論文で提案されている体系的な SPARQL の意味論と、それに基づく複雑性、正規化、最適化の形式的な考察が、従来研究に比べた本論文の貢献である。

2.2 研究の内容

本論文の2節では、SPARQL の構文規則とそれに基づく構成的な意味論の定式化について説明がなされる。基本となる表現は RDF triple, $(subject, predicate, object) \in (IUB) \times I \times (IUBUL)$ である。ここで、 I, B, L はそれぞれ IRIs, Blank nodes, literals であり、互い

*1 「スパークル」と読む。正式名称は「SPARQL Protocol and RDF Query Language」である。

に素な集合である。\$T\$は、和集合 \$I \cup B \cup L\$ であり、変数の集合は \$V\$ とする。本論文では、これら集合に対して UNION, AND, OPT, FILTER の2項演算子を利用することで SPARQL のグラフパターンの構文規則を表現している (2.1 節)。

この構文規則に基づき、\$V\$ から \$T\$ への写像関数 \$\mu: V \rightarrow T\$ を考える。このとき本論文では SPARQL グラフパターンの意味論を、パターン表現 \$P\$ を入力とし、写像 \$\mu\$ の集合を出力とする関数 \$[P]_\mu\$ として定義する。つまり RDF のデータ集合 \$D\$ にマッチする写像 \$\mu\$ の集合をもって意味論とする [C. Gutierrez 04]。詳細は論文の 2.2 節を参照されたい。提案された意味論に基づく SPARQL グラフパターンの表現と処理の具体的な実例が示されている。

本論文では続いてクエリ一言語としての SPARQL の複雑性について考察を行っている (3 節)。SPARQL のグラフパターンの意味論に基づき、本論文では、入力 RDF データ集合 \$D\$、グラフパターン \$P\$、写像関数 \$\mu\$ が与えられたときに、\$\mu \in [P]_\mu\$ を満たすかどうかという問題 (EVALUATION) についての評価に基づき複雑性の考察を行っている。SPARQL の複雑性に関して次のような定理が成り立つことが示されている。

[定理 1] AND および FILTER 演算子のみに基づく SPARQL グラフパターン表現について、EVALUATION を \$O(|P|, |D|)\$ 時間で解くことが可能である。

[定理 2] AND, FILTER および UNION 演算子のみに基づく SPARQL グラフパターン表現について、EVALUATION は NP 完全である。

[定理 3] SPARQL グラフパターン表現について、EVALUATION は多項式領域 (PSPACE) 完全である。

[定理 4] グラフパターン表現 \$P\$ が与えられたとき、\$P\$ に対する評価問題 EVALUATION(\$P\$) は、すべてのグラフパターン表現 \$P\$ について対数領域 (LOGSPACE) である。

複雑性の考察に続いて、4 節以降では従来のアプローチに対する比較を行っている。本論文で提案している構成的な意味論に対して、従来は SPARQL のグラフパターン表現に対して手続き的な意味論が適用されてきた。つまり、グラフパターンの解析木に対して深さ優先探索を行うわけである [arq]。

従来のアプローチは、それが構成的でないため、例えば、複数のデータ集合からの結果を統合する際や、クエリの結果を再利用する際に問題となる。もし、本論文で提案する構成的なアプローチと従来のアプローチが与える結果が、すべての RDF データセットに対するクエリについて一致するのであれば、構成的なアプローチが形式的意味論を与える一方で、深さ優先探索によりクエリの効率的な処理が可能である。しかしながら、必ずしもすべてのグラフパターンにおいて、これらのアプローチが一致するわけではない。そこで、本論文では、

これらのアプローチがどのような条件のもとで一致するかについて考察を行っている。そして、グラフパターンが “well-designed” であるときに、従来の操作手続き的な意味論と本論文で提案する意味論が等価であることを示している (4.2 節定理 5)。ここで、グラフパターン \$P\$ が well-designed であるとは、\$P\$ のすべてのサブパターン \$P' = (P_1 OPT P_2)\$ と \$P\$ のすべての変数 \$?X\$ に対して、「もし \$?X\$ が \$P_2\$ の内部および \$P'\$ の外部に含まれるとき、\$?X\$ は \$P_1\$ にも含まれる」という OPT 演算子に関する条件として定義される。

最後に SPARQL の正規化について、すべての well-designed なグラフパターン \$P\$ が

$$(\dots (t_1 AND \dots AND t_n) OPT O_1) \dots) OPT O_n$$

で表される正規形におけるパターンと等価であることを示している。ここで、\$t_i\$ はトリプル、\$O_j\$ は同正規形を表す。これに基づき SPARQL の最適化手続きについて議論を行っている (4.3 節)。

2.3 研究の意義

本論文の貢献は、SPARQL クエリの基本となるグラフパターンマッチについて、その構成的な意味論を提案している点である。そして、提案に基づき、従来研究では明らかにされてなかった SPARQL のグラフパターンの評価について、その複雑性が PSPACE 完全であることを示している。また、グラフパターンの標準形についての考察と、それに基づく最適化手続きは、SPARQL を利用する開発者、ユーザに大きな示唆を与えている。さらに、本論文では、従来の SPARQL に用いられてきた手続き的な意味論との比較を行い、特定の条件のもとでそれらが同一になることを示している。これは、SPARQL グラフパターンについて明確な意味論を保持しながら効率的な処理を実現する際に有益となる考察である。

3. Wolf, S., Jeff, Z., and Uwe, T.: Querying the Semantic Web with Preferences

3.1 研究の背景と目的

本論文 [Wolf 06] では Preference を用いてデータにクエリを行うための SPARQL の拡張手法と具体的な実装について述べている。本論文は、特にセマンティック Web における検索やサービスマッチングなどに利用可能な SPARQL の応用研究である。

データ (例えば RDF のデータ集合) に対して検索を行う際、対象データが膨大であるほど、検索クエリの結果もまた膨大なものになる。そのため、多くの場合は検索結果には何らかのスコアリングにより完全な順位付けがなされる。しかしながら、ユーザが複数の評価指標で結果をとらえようとする場合、完全な順位付けは不十分である。本論文から引用すると、例えばセラピストを探しているときに、ユーザはセラピストが近くに

るのか離れているのか、評判がどれくらい良いか、予約の時間はいつが良いか、など検索において複数の指標をもっている。Preferenceは、このような複数の評価指標を柔軟に表現するのに用いられる概念である。先の例において、ユーザは例えば、評判がなるべくよい、混雑する時間帯を避けたい、午前よりは午後がよい、などのPreferenceをもつだろう。

ユーザのPreferenceをクエリーとして処理するには、ユーザがどのような結果を“好む”かに応じて、そのPreferenceをソフト制約としてクエリーに付加すればよい。このとき、クエリーの処理は制約に基づき支配的(dominated)でない解をもとめる制約充足問題としてとらえることができる。

3.2 研究の内容

提案手法の基本的な考え方は、Chomickiが提案している関係データベースにおけるPreferenceクエリーの定式化[Chomicki 03]に基づくものであり、それらをSPARQLのもとで拡張を行ったものである。

具体的には、SPARQLクエリーの結果に対してさまざまな操作を与えるSolution modifiersに拡張を行いPreferenceを処理可能なmodifierの設計と実装を行っている。拡張においては、一般にPreferenceを処理する際に考慮される、次のような点が特徴である。第1点として、ほかのいかなる解からも支配的でないすべての解を返すということ。第2点として、すべてのハードおよびソフトな制約を充足する解を返すということである。なお、制約を充足する解がない場合は、一部またはすべてのソフト制約をゆるめたとうえで最適解のみを返す。これらの実装はARQを拡張することでなされている。なお、Preferenceについての構文規則や意味論についての詳細は本論文を参照されたい。

3.3 研究の意義

セマンティックWebデータの検索における現在のアプローチのほとんどは、限定的なスコアリングや順序付けの機能を提供するのみであり、順位付けやPreferenceを形式的に表現可能なクエリー言語はない。

本論文では、関係データベースにおいて確立されたPreferenceを用いたクエリーの概念がセマンティックWebのクエリーにおいても同様に重要な役割を果たすことを示しており、今後セマンティックWebデータの検索において非常に有用となる手法である。

4. ま と め

メタデータのアノテーションの半自動化、エンドユーザアプリケーションの普及などにより、セマンティック

Webのデータは徐々に蓄積されてきている。それに伴い、RDFやOWLのようなデータの表現に加えて、データに対してどのようにクエリーを出すかが現在、セマンティックWebの研究の焦点の一つになってきている。SPARQLはRDFクエリー言語としてセマンティックWebのクエリー言語の基盤を与えるものである。

本稿では、ISWC2006からSPARQLに関連した論文を紹介し、研究の動向について概観を行った。SPARQLを中心としたセマンティックWebにおけるクエリー言語は、セマンティックWebアプリケーションを構築する際に必須のものであり、研究の面のみならず実用の面においても今後、重要となるであろう。ISWC2006のWebページではメタデータ化された学会、論文データに対してSPARQLを用いて検索を行える機能を提供している*2。本稿で紹介した論文に加えて、実際のデータに触れながらSPARQLに対する理解を深めていただければ、幸いである。

◇ 参 考 文 献 ◇

- [Arq] ARQ-A SPARQL Processor for Jena, version 1.3 March 2006, Hewlett-Packard Development Company.
- [C. Gutierrez 04] C. Gutierrez, C. H. and Mendelzon, A.: Foundations of semantic web databases, *Proc. PODS* (2004)
- [Chomicki 03] Chomicki, J.: Preference formulas in relational queries, *ACM Trans. Database Syst.*, Vol. 28, pp. 427-466 (2003)
- [Jorge 06] Jorge, P., Marcelo, A. and Claudio, G.: Semantics and complexity of SPARQL, *Proc. 5th Int. Semantic Web Conference (ISWC2006)*, pp. 30-43 (2006)
- [spa] SPARQL Query Language for RDF, W3C Candidate Rec. 6 (April 2006)
- [Wolf 06] Wolf, S., Jeff, Z. and Uwe, T.: Querying the semantic web with preferences, *Proc. 5th Int. Semantic Web Conference (ISWC2006)*, pp. 612-624 (2006)

2007年3月4日 受理

—— 著 者 紹 介 ——



森 純一郎 (学生会員)

2001年東北大学工学部情報工学科卒業。2003年東京大学大学院情報理工学系研究科修士課程修了。現在、同大学院博士課程在学中。2006年より客員研究員としてドイツ人工知能研究所(DFKI)においてユビキタス環境における情報共有の研究を行っている。ユーザモデリング、セマンティックWeb、Webマイニング、情報抽出などの研究に興味をもつ。

*2 <http://iswc2006.semanticweb.org/>