

オントロジーマッピングを用いた意味情報に基づく オープンデータへのアクセス効率化機構の試作

Toward a Better Access Mechanism to Open Data by using Ontology Mappings and their Underlying Semantics—A Preliminary Report

福田 直樹*¹

Naoki FUKUTA

*¹静岡大学 学術院情報学領域

Department of Informatics, Shizuoka University

In this paper, we report our preliminary implementation of a set of software frameworks and tools to make open data accesses efficient and easy, based on our ontology mapping-based query transformation approaches and some AI-based techniques. The implemented techniques realize more efficient accesses to open data despite the use of indirect access from mapped ontologies, as well as efficient and easier retrieval of open data sources that could be used for an application.

1. はじめに

これまでに著者およびそのグループは、オントロジーマッピングの効果的な利用技術についての検討を行ってきている。

たとえば、SPARQLoid[Fujino 12a][Fujino 12b][Fujino 14]では、オープンデータへのアクセスで用いられる SPARQL エンドポイントへのアクセスを、そのオープンデータそのものに対してのオントロジーに必ずしも熟知しない場合であっても、他のオントロジーからのマッピングを用いてアクセス可能とすると同時に、その際のマッピング精度に基づく取得データの順序付けを、クエリ書き換え技術により実現している [Fujino 12a]。また、SPARQLoid ではさらにそのアクセスを 1 つのクエリから複数のエンドポイントにまたがった検索を行う Federated Query に拡張 [Fujino 12b] しており、その有効性についての検証 [Fujino 14] を行っている。

また、オープンデータのエンドポイントそのものの探索問題 [Ladwig 10] に対しては、たとえば、クエリ作成時にそのクエリの実行に適したエンドポイントを、その検索対象に対する文字列や既存オントロジーとのマッピングに基づいてエンドポイントの適合性を探索しながらそのクエリの実行を可能とする機構 [Noguchi 13] の開発を進めてきている。

こうしたオープンデータへのアクセス方法の拡張やエンドポイント探索手法の実装では、それらのエンドポイントへのアクセス手法の効率化を行わないと、エンドポイントやその間の通信路に大きな負荷がかかってしまう点が課題となる [Kadono 14b]。この課題に対する 1 つのアプローチが、こうした探索問題をオンライン学習手法問題の 1 つである BLMAB(Budget-Limited Multi-Armed Bandit) 問題として定式化・拡張 [Kadono 14b] し、BLMAB アルゴリズムやその派生アルゴリズム等 [Kadono 14a] を用いて効率化を試みる方法である。

分散データソースへのアクセス効率化手法としては、これらのアクセス時のクエリやアクセス先選択の効率以外にも、データソースとの間のネットワーク構成上の非対称性などを利用したアクセスの効率化手法があり、たとえば、P2P 型・モバイルエージェント [White 94] 型のアクセスの効率化手法が提案されてきている [Fukuta 12]。本研究の目的は、これらの複数

のデータアクセス手法の効率化手法を、オープンデータへのアクセス方法の容易化・拡張手法と組み合わせて利用可能とすることで、効率的で容易なオープンデータへのアクセスを実現するためのソフトウェア基盤の実現である。本研究では、その実現のコアとなる機構の試作について述べる。

2. 試作機構の構成と概観

本機構は、現時点では、モビリティを持たせたソフトウェアの開発実行プラットフォームである MiLog[Fukuta 01] を用いて試作を進めている。MiLog は、文献 [Fukuta 12] 等での実績があり、本試作における実装・検証の効率化には有効であると考えられる*¹。

図 1 は、本システムの動作の概観を示すものである。図 1 では、SPARQL 拡張アクセス処理エンジンを、データソース(右下)、クライアントサイド(左上)、およびその中間となる proxy サイト(右上)となるホスト上にそれぞれ本実行環境を起動している。この例では、各ホストはそれぞれ OS X 10.10, Ubuntu Linux 14.04(仮想環境上)、および Ubuntu Linux 10.04(仮想環境上)で動作しており、それぞれの動作環境上で MiLog を動作させ、本試作システム上で試作した機構を動作させている。ここでの仮想環境の実行には、Parallels 8.0 を用いている。各ホスト上での SPARQL 処理エンジンは、この例では仮に MiLog 上で簡易的に準備したものをしている。

このように、必要に応じて処理エンジン等を異なるホスト上に配置可能とし、その処理性能および負荷等の観測を行えるようにしている。図 1 では、テストクエリを実行した際のクエリの動作の流れをモニタする状況を示している。

これら以外にも、MiLog の持つ Web サーバ機能を用いて、クライアントサイドでの Web ブラウザ経由での簡易ユーザインタフェース等との連携機能の試作を進めている。

3. おわりに

本研究では、効果的にオープンデータにアクセスし、意味情報に基づいて適切な情報を検索・提示するためのソフトウェア

*¹ 一方で、より広い一般からの本機構の利用時には、より普及の進んだ言語やプラットフォーム上で動作するものの方が好まれると考えられる。そうした実装は、試作による知見に基づいて、別途準備することを考えている。

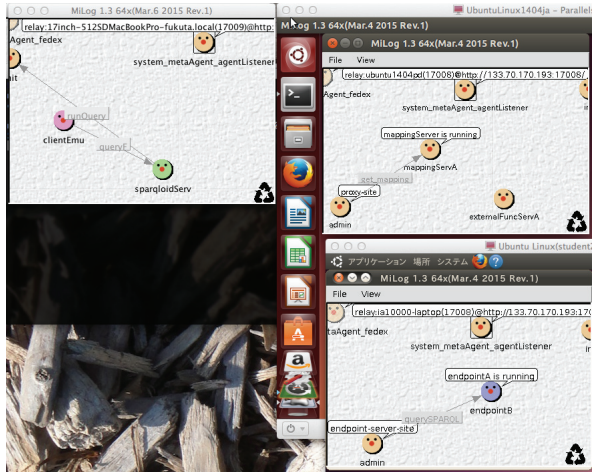


図 1: 試作システムの概観 (クエリ実行モニタ)

基盤におけるアクセス効率化機構の試作についての現状と課題を述べた。本ソフトウェア基盤および試作した機構の特長は、オントロジーマッピングに基づく異種データ検索技術にクエリ変換技術およびオンライン学習技術などを適用することで、必要となるデータの提供源そのものをも、種々のオープンデータから動的にかつ効率的に検索可能とする技術の実現を行おうとしている点である。

本稿ではそのアクセス効率化機構の試作の概要のみを述べたが、その性能の詳細な解析、およびこれらを応用したシステムの実装事例における利点の検証とその知見の考察は、今後の課題である。

また、SPARQL クエリ実行の効率化手段としては、RDFS 等の構造を利用したクエリの書き換え [Bischof 13] など、様々なアプローチが試みられている。著者およびそのグループでも、エンドポイント上での OWL 推論を可能とした場合における実行時間の増加に対処する手法の開発 [Yamagata 14b][Yamagata 14a] を行っており、これらと本機構の統合的な利用を行えるようにすることも、今後の課題である。

謝辞

本研究の一部は、電気通信普及振興財団研究助成の支援による。

参考文献

- [Bischof 13] Bischof, S. and Pollers, A.: RDFS with Attribute Equations via SPARQL Rewriting, in *Proc. the 10th Extended Semantic Web Conference (ESWC2013)*, pp. 335–350 (2013)
- [Fujino 12a] Fujino, T. and Fukuta, N.: A SPARQL Query Rewriting Approach on Heterogeneous Ontologies with Mapping Reliability, in *Proc. IIAI International Conference on Advanced Applied Informatics (IIAI-AAI 2012)*, pp. 230–235, Fukuoka, Japan (2012)
- [Fujino 12b] Fujino, T. and Fukuta, N.: SPARQLoid - a Querying System using Own Ontology and Ontology Mappings with Reliability, in *Posters and Demonstra-*

tions Track, The 11th International Semantic Web Conference (ISWC2012) (2012), (demonstration)

- [Fujino 14] Fujino, T. and Fukuta, N.: Utilizing Weighted Ontology Mappings on Federated SPARQL Querying, in Kim, W., Ding, Y., and Kim, H.-G. eds., *Lecture Notes in Computer Science*, Vol. 8388, pp. 331–347, Springer-Verlag (2014)
- [Fukuta 01] Fukuta, N., Ito, T., and Shintani, T.: A Logic-based Framework for Mobile Intelligent Information Agents, in *Poster Proc. of the Tenth International World Wide Web Conference (WWW10)*, pp. 58–59 (2001)
- [Fukuta 12] Fukuta, N.: A Mobile Agent Approach for P2P-based Semantic File Retrieval, *Journal of Information Processing*, Vol. 20, No. 3, pp. 607–613 (2012)
- [Kadono 14a] Kadono, Y. and Fukuta, N.: LAKUBE: An Improved Multi-armed Bandit Algorithm for Strongly Budget-Constrained Conditions on Collecting Large-Scale Sensor Network Data, in *Proc. 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI2014)*, pp. 1089–1095 (2014)
- [Kadono 14b] Kadono, Y. and Fukuta, N.: An Online Learning-based Efficient Search System for Sufficient SPARQL Endpoints using Extended Multi-armed Bandit Algorithm, in *Poster and Demo Proc. of the 4th Joint International Semantic Technology Conference (JIST2014)* (2014), (poster with demonstration)
- [Ladwig 10] Ladwig, G. and Tran, T.: Linked Data Query Processing Strategies, in *Proc. International Semantic Web Conference (ISWC2010) PART I*, pp. 453–469 (2010)
- [Noguchi 13] Noguchi, H., Fujino, T., and Fukuta, N.: On Implementing SPARQLoid and its Query Coding Support Framework – Querying with Weighted Ontology Mappings, in *Proc. The 3rd Joint International Semantic Technology Conference (JIST2013)* (2013), (demonstration)
- [White 94] White, J. E.: Mobile Agents Make a Network an Open Platform for Third-Party Developers, *IEEE Computer*, Vol. 27, No. 11, pp. 89–90 (1994)
- [Yamagata 14a] Yamagata, Y. and Fukuta, N.: Approximating Inference-enabled Federated SPARQL Queries on Multiple Endpoints, in *Proc. ISWC2014 Posters and Demonstrations Track, a track within the 13th International Semantic Web Conference (ISWC2014)*, pp. 441–444 (2014)
- [Yamagata 14b] Yamagata, Y. and Fukuta, N.: A Dynamic Query Optimization on a SPARQL Endpoint by Approximate Inference Processing, in *Proc. 3rd IIAI International Conference on Advanced Applied Informatics (IIAI AAI2014)*, pp. 161–166 (2014)