

LOD (Linked Open Data) による 地域情報の発信と活用に関する研究

斎藤 一 長尾 光悦 諸岡 卓真
北海道情報大学

A Study of Application of Linked Open Data in Local Community

Hajime SAITO, Mitsuyoshi NAGAO and Takuma MOROOKA
Hokkaido Information University

平成27年11月

北海道情報大学紀要 第27巻 第1号別刷

〈報 告〉

LOD(Linked Open Data)**による地域情報の発信と活用に関する研究**

齋藤 一* 長尾光悦† 諸岡卓真‡

**A Study of Application of Linked Open Data
in Local Community**

Hajime Saito* Mitsuyoshi Nagao† Takuma Morooka‡

要旨

近年、政府や自治体が公共のデータをオープンデータとして公開する等、世界的にオープンデータ化が進んできている。また、政府や自治体に限らず、個人や企業も様々なデータをオープンデータとして公開している。一方、それに合わせて公開されたオープンデータを連携・活用するための技術が必要になってきている。オープンデータを活用する技術として、セマンティック Web とよばれるコンピュータが自律的に情報を処理・利用する技術があり、その技術のひとつに Linked Open Data (LOD) がある。本稿では、平成 26 年度学内共同研究として実施した、LOD による地域情報の発信と活用に関する研究について、報告を行う。本研究では、LOD の利用・活用方法の調査および地域情報の発信を行うために、北海道の蝶のデータと観光スポットのデータを LOD 化した。また、LOD を利用するための環境である SPARQL エンドポイントおよびサーバの構築を行った。更に LOD 間で情報を連携し、関連する情報を取得・表示するアプリケーションの開発を行い、動作確認を行った。

キーワード

LOD 観光情報 SPARQL

*情報メディア学部情報メディア学科 准教授, Associate Professor, Department of Information Media, Faculty of Information Media

†経営情報学部システム情報学科准教授, Associate Professor, Dept. of Systems and Informatics

‡経営情報学部先端経営学科准教授, Associate Professor, Dept. of Business and Information Systems

1. はじめに

データや文書等の著作権が持つ独占的な権利を緩め、誰もが自由に利用することができ扱いやすい公共財としてデータを増やす「オープンデータ」の運動が世界的に広がっている[1]。オープンデータは単にデータを公開することではなく、二次利用や商用利用が認められた形でのデータ提供を指す。Linked Open Data (以下, LOD) [2][3][4]は, Web の情報を意味的に関連づけて, 直接的に関連する情報を辿れるようにする技術である。World Wide Web(WWW)の考案者であるティム・バーナーズ=リーは LOD の原則[5]を以下のように定義している。

- あらゆるデータの識別子として URI を使用する,
- 識別子には HTTP URI を使用し, 参照やアクセスを可能にする,
- URI にアクセスされた際には有用な情報を標準的なフォーマット (RDF 等) で提供する,
- データには他の情報源における関連情報へのリンクを含め, ウェブ上の情報発見を支援する。

LOD を活用した地域情報の配信は, 横浜のアートに関する情報を LOD 化したヨコハマ・アート・LOD プロジェクト[6]や, 函館まちあるきマップの観光ナビゲーション[7]等, 国内でも幾つか事例がある。本研究では, これらの事例にはない LOD を活用したコンテンツの試作として, 下記の 3 つを実施し, 本学における LOD の利活用を押し進める。以下では, 各項目の成果について説明する。

1. LOD の配信技術 (SPARQL End Point) に基づいたサーバ構築
2. 公開可能なデータの整理と配信
3. LOD を活用したコンテンツの試作



図 1 RDF の構造

2. 実施内容

2.1 LOD 配信技術に基づいたサーバの構築

LOD は, RDF (Resource Description Framework) [8]形式でデータの発信を行う。RDF は, コンピュータが情報を扱いやすくするメタデータの表現形式である。RDF は, 英語の構文に基づく情報の表現方法であり, データを主語・述語・目的語の 3 要素 (トリプル) で捉え, それぞれの関係性を記述する枠組みである。例えば, 主語[吾輩は猫である]と, 目的語[夏目漱石]の関係を, 述語[著者]を構造化して表現する (図 1)。

公開したい情報を LOD として発信することで, 異なるデータベース上のデータを相互にリンク・共有することができる。これにより, Web 全体を巨大な知識データベースとして利用することが可能となる。RDF 情報を入出力するためには, SPARQL End Point を用いる。SPARQL End Point とは, クエリ言語である SPARQL[8]により RDF の検索・取得を行う WebAPI である。

本研究では, 以下の二つの方法で, SPARQL End Point を実装した。

(1) Fuseki [9]

(2) Google Spreadsheet [10]

Fuseki は, Apache Jena プロジェクトによって提供された SPARQL エンドポイントを構築するためのオープンソースソフトウェアである。また, TDB というデータベースシステムを同梱しており, RDF を格納するための RDF ストアにも対応しているため, SPARQL End Point 構築と合わせて RDF 用の DB も構築することができる。

Google Spreadsheet を用いた方では, Google JS API, Google Visualization API,

Google Feed API, Google App Script (JavaScript) により, 検索や表示の機能を開発した。また, 単に Excel ファイルを Web サーバで公開しても, LOD と呼ぶこともできるようであるが, 本研究では, 上述した LOD の原則を意識し, 第三者が利用可能な形での公開を前提とするため, SPARQL End Point の実装を重視した。

2.2 公開可能なデータの整理と配信

LOD として, 以下の 3 種類のデータを用意した。以下にその詳細を示す。

- ① 北海道の蝶[11]
- ② 北海道の観光スポット[12]
- ③ 東北に関する情報 (東北ブレイブズ)

2.2.1 北海道の蝶

北海道の蝶に関しては, 北海道博物館 (旧北海道開拓記念館) に調査に行くことで, 国内, 特に道内には蝶に関する Web[13][14]が少ないことが分かり, 本研究で蝶を取り上げることにした。図書[15][16]を参考に, 15 のフィールド (ID, 名称, 学名, 見られる時期, 大きさ, 見られる場所, 食べるもの, 模様, 色, 絶滅具合, 画像, URL, 上科, 科, 亜科) を設けた。フィールド名を作成する際には, "namespace lookup for RDF developers [17]" というサイトを利用し, 既存のフィールド名をできる限り使用することで, 学外の LOD とも連携しやすくなるようにした。現時点で 127 件のデータを登録し, Fuseki を用いた SPARQL End Point で公開している [11][18]。

2.2.2 北海道の観光スポット

北海道の観光スポットとして, 千石涼太郎の著書「なまら北海道だべさ!!」[19]を参考に, 温泉, 会社, 遊び, 食べ物, 北海道遺産, 重要文化財, 風習, 時節, イベント, 日用品, ゆるきゃらの 11 のカテゴリに分けた。フィールド名としては, 名称, 場所, 備考, ポイント (観光のみどころ), 概要, 歴史, 関係性, リファレンスの 8 つを設けた[20]。LOD とし

ては, 402 件のデータを, Fuseki を用いた SPARQL End Point にて公開している[12]。また, 北海道の観光スポットは, Fuseki と Google Spreadsheet の二つの方法で公開しており, 後者は「オンリー北海道 Lab プロジェクト」としてサイトも公開している[21]。



図 2 オンリー北海道 Lab プロジェクト

2.2.3 東北に関する情報 (東北ブレイブズ)

学生プロジェクトで制作した, 東北に対する関心を高めることを目的としたカードゲーム「東北ブレイブズ[22]」について, そのカード制作に用いたデータを LOD 化した。フィールドとして, カード名, 読み方, 地域, カテゴリ, URL, 種別, アピールポイント, カードの効果, カードの元ネタ, 効果の設定理由, SVG の 11 を設定し, 現在 41 件のデータを用意, 公開の準備を進めている。この 41 件には, 北海道のカードと東北のカードを対戦させた方が東北への関心が高まるとの考えから, 北海道のデータも含まれている。

2.3 LOD を活用したコンテンツの試作

上述した北海道蝶と観光スポット 2 つの SPARQL End Point を利用した情報連携アプリケーションの開発を行った。問い合わせにより取得した情報をもう一方のエンドポイントの問い合わせに利用することで, 関連性の

ある情報を取得し、結果をページに表示することができる(図3)。SPARQL End Point用サーバはDell社製のPowerEdge T320タワーサーバ(表1)を使用した。また、2つのエンドポイントを構築するため仮想化を行った。仮想化は仮想化ソフトであるVMware vSphere Hypervisor ESXi 5.5を使用した。OSはCentOS7.0を使用した。これにより、Fusekiサーバを構築し、前述した蝶と観光スポットのRDFデータを各サーバのRDFストアに追加した。動作確認では、最初に観光スポットLODに問い合わせを行い、見たい名所を決定する。決めた名所の季節の情報を問い合わせし(図4)、結果として、観光スポットの情報に加えて、そのスポットで観測できる蝶の情報を取得できることを確認した(図5)。

表1 サーバのスペック

Dell社製 PowerEdge T320 タワーサーバ	
CPU	インテル® Pentium® プロセッサー 1403 v2 (2.60GHz, 6Mキャッシュ, 2C, 80W) Max Mem 1333MHz
メモリ	8GB RDIMM, 1600MT/s, LV, 2R x8 Data Width
ハードディスク	500GB 7,200RPM (SATA II HDD/3.5インチ/ホットプラグ非対応)
電源ユニット	シングル, ケーブル式電源ユニット, 350W



図3 LOD 連携アプリケーション

```

問い合わせ文(オンリー北海道)
prefix schema: <http://schema.org/>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
select ?名所・名物 ?見所 ?季節 where {
  ?code rdfs:label ?名所・名物 ;
  schema:season ?季節 ;
  schema:lookpoint ?見所 ;
  schema:category ?カテゴリー ;
  filter regex (?名所・名物, "十勝岳温泉") .
} limit 10
    
```

図4 システムへの問合せ

名所・名物	見所	季節
十勝岳温泉	北海道で最も高所の温泉地でもあり、登山基地としても利用される。十勝岳の標高約1,290mの位置にある湯室間は、条件がそろえば露天風呂から雲を見下ろすことができる。酸性・鉄・硫酸塩泉。カルシウム・ナトリウム-硫酸塩泉	夏・秋

見ることができる蝶(現在は問い合わせ結果の一番上の名所・名物の季節)
※春3~5月、夏6~8月、秋9~11月、冬12~2月

名称	学名	時期	場所	食べるもの	色	絶滅具合
ウスバキチョウ	Parnassius eversmanni	6月,7月	標高1,700m以上の高山帯のみ	コマクサ	黒と黄色	★3
ヒメウスバシロチョウ	Parnassius stubendorfi	6月	道内に広く分布し、山道や渓流沿いで見られる	エゾエンゴサク	白	★2
ウスバシロチョウ	Parnassius citrinarius	6月	道内では渡島半島、石狩低地帯から日高山脈、帯広周辺、根釧台地などに分布している	エゾエンゴサク、ムラサキケマン	白	★2
アゲハ	Papilio xuthus	3月4月, 5月6月, 7月8月	道内に広く分布している	キハダ、サンショウ	黒と黄色	★1
キアゲハ	Papilio machaon	3月4月, 5月6月, 7月8月	道内に広く分布している	シシウド、セリ、ボウフウ、イワミツバ	黒と黄色	★1
オナガアゲハ	Papilio macilentus	5月,7月	北海道西部~南部に分布している	コクサギ、サンショウ、カラスアゲハ	黒	★1

図5 統合情報の出力

3. まとめと今後の課題

本研究では、LODを用いた地域情報の発信と活用のために、SPARQL End Pointを構築、北海道の蝶、観光スポットの2つの情報を公開した。また、本学では我々以外にも地域コンテンツに関して研究をしている教員が多い。LODを本学に導入することで、本学の研究成果を学外の方々にも活用してもらえらる可能性が高まると期待する。

参考文献

[1] 庄司昌彦他, 特集オープンデータの活用, 情報処理, Vol.54, No.12, pp.1201-1247, 2013.

- [2] 伊藤健太郎, 佐藤勇紀, 濱崎俊, ためしてわかるセマンティック Web ～次世代型データ活用術～, 技術評論社, 東京, 2007.
- [3] 武田英明, “オープンデータと Linked Open Data”
http://www.meti.go.jp/committee/kenkyukai/shoujo/it_yugo_forum_data_wg/pdf/003_03_00.pdf. (参照: Jan, 22, 2015).
- [4] Michael, “5 ★ オープンデータ”
<http://5stardata.info/ja/>
 (参照: Jan, 21, 2015).
- [5] Tim Berners-Lee, “Linked Data – Design Issues”,
<http://www.w3.org/DesignIssues/LinkedData.html>. (参照: Jan, 19, 2015).
- [6] ヨコハマ・アート・LOD プロジェクト,
<http://ocdi.jp> (参照: Feb. 6, 2014)
- [7] 奥野 拓, Linked Open Data を用いた地域観光情報コンテンツ活用の試み, 観光情報学会 第8回研究発表会講演論文集, pp.53-56, 2013.
- [8] Cyber Librarian, “RDF 用クエリ言語 SPARQL”,
<http://www.asahi-net.or.jp/~ax2s-kmt/internet/rdf/rdf-sparql-query.html>
 (参照: Jan, 26, 2015).
- [9] Fuseki: serving RDF data over HTTP,
http://jena.apache.org/documentation/serving_data/ (参照: Aug. 31, 2015).
- [10] Google Spreadsheet,
https://www.google.com/intl/ja_jp/sheets/about/ (参照: Aug. 31, 2015).
- [11] 北海道の蝶 SPARQL End Point,
<http://hiu-lod3.do-johodai.ac.jp/>
 (参照: Apr. 10, 2015)
- [12] 北海道の観光スポット SPARQL End Point, <http://hiu-lod2.do-johodai.ac.jp/>
 (参照: Apr. 10, 2015)
- [13] ぷてろんワールド・蝶の百科ホームページ, <http://www.pteron-world.com/>
 (参照: Apr. 10, 2015)
- [14] 国立科学博物館「MOTHPROG」,
<http://www.mothprog.com/>
 (参照: Apr. 10, 2015)
- [15] 木野田君公, 札幌の昆虫, 北海道大学図書刊行会, 2006.
- [16] 日本チョウ類保全協会, フィールドガイド 日本のチョウ, 誠文堂新光社, 2012.
- [17] namespace lookup for RDF developers,
<http://prefix.cc> (参照: Apr. 10, 2015)
- [18] 村上廉斗, 諸岡卓真, 斎藤一, 長尾光悦, Linked Open Data による地域情報の発信と活用,
 2015 年電子情報通信学会総合大会・ISS 特別企画「学生ポスターセッション」, ISS-P-63, 2015.
- [19] 千石涼太郎, なまら北海道だべさ!!, 北海道新聞社, 2007.
- [20] 齊藤成美, 斎藤一, 長尾光悦, 諸岡卓真, 隼田尚彦, 向田茂, 安田光孝, 北海道の独自性を学ぶための LOD を活用した Web サイトに関する研究 ～有名観光コンテンツと似て否なるデータを表示させるシステムの提案～, 観光情報学会第 9 回研究発表会講演論文集, pp.30-31 2014.
- [21] オンリー北海道 Lab プロジェクト,
<http://sherry.do-johodai.ac.jp/doc/>
 (参照: Apr. 10, 2015)
- [22] 佐藤雄哉, 嶋津貴之, 佐藤佑起, 吉野正樹, 斎藤一, 東北を知るきっかけを作るカードゲーム「東北ブレイブズ」の開発, 観光情報学会第 11 回全国大会講演論文集, pp.20-21, 2104.

謝辞

本研究は平成 26 年度学内共同研究『LOD (Linked Open Data) による地域情報の発信と活用に関する研究』により行われました。