



私のブックマーク

「Linked Open Data」^{†1}

加藤 文彦 (情報・システム研究機構)

1. はじめに

Linked Open Data (リンクトオープンデータ, LOD) は, 再利用しやすいデータを Web 上で公開・共有するための技術や方法論の総称です. 主にセマンティック Web で培われてきた技術などが適用されているため, 研究コミュニティとしてはセマンティック Web の一分野という扱いになっています.

Linked Open Data の概要を知りたい場合には, 以下のビデオ資料がおすすめです.

- **Linked Open Data - What is it?**

Linked Open Data とは何かを芸術文化データを例にわかりやすく説明しています.

- **ティム・バーナーズ＝リーが示す次の Web**

World Web Wide の発明者である **Tim Berners-Lee** 卿の **TED 2009** での講演です.

- **オープンデータとマッシュアップで変わる世界**

Tim Berners-Lee 卿の **TED 2010** での講演です.

2. Open Data

Linked Open Data は大きく分けて **Open Data** と **Linked Data** という二つの動向が混ざり合っているため, 双方のことを知っておく必要があります.

再利用しやすいデータを考えるうえで, 重要な要素の一つはライセンスです. **Open Data** はデータを配布する際に一定のライセンスを適用することによって, データの消費者が自由にアクセスや再利用, 再配布をすることを可能とします. **Open Data** に関連するリンクは以下のとおりです.

- **The Open Definition**

データやコンテンツにとっての **Open** とは何かを定義している文書. 定義に適合するライセンスのリストも公開しています.

- **Creative Commons**

Open Data のライセンスとしてよく使用されている **Creative Commons** ライセンスを提供しています.

- **Open Data Commons**

Open Data のライセンスとしてよく使用されています.

- **Open Data Index**

Open Data の一指標

- **the Datahub**

世界中のデータカタログをつくることを目的としたサイトです.

- **CKAN**

データポータルサイトをつくるためのソフトウェアです.

- **Socrata**

データポータルソフトウェアの一つです.

- **Open Data Directory**

Open Data の作成や公開に関わっている組織を登録できるサイトです.

^{†1} http://www.ai-gakkai.or.jp/my-bookmark_vol28-no5

- CKAN 日本語

日本国内のデータカタログをつくることを目的としたコミュニティプロジェクトです。

行政・公共データ

Open Data で最も大きな割合を占めているのが行政や公共のデータになります。現在世界中の政府や地方自治体などがデータ公開のためのポータルサイトを立ち上げるようになっていきます。

- Data.gov

アメリカ政府のデータポータル

- Data.gov.uk

イギリス政府のデータポータル

- European Union Open Data Portal

ヨーロッパ連合のデータポータル

- Europe's Public Data

ヨーロッパの公共データの横断データポータル

- Open DATA METI

経済産業省のデータポータル

- データシティ鯖江

日本で初めて Open Data という形でデータを公開し始めた福井県鯖江市のサイト

- オープンデータ流通推進コンソーシアム

総務省が行っているオープンデータを推進するためのコンソーシアム

3. Linked Data

Linked Data は構造化データを適切に公開・共有するための Web 技術です。Linked Data は World Wide Web のアーキテクチャを踏襲して設計されていますので、Web アーキテクチャー一般に関するリンクも含みました。

- Information Management: A Proposal

Tim Berners-Lee 卿による World Wide Web の提案文書。すべてはここから。Linked Data の基礎概念もここにすでに入っています。

- Architecture of World Wide Web, Volume One

World Wide Web のアーキテクチャを一通り勉強するための資料。W3C 勧告。

- Web Architecture from 50,000 feet

Tim Berners-Lee 卿による Web アーキテクチャの俯瞰。

- Architectural Styles and the Design of Network-based Software Architectures

Roy T. Fielding 氏による REST の論文です。REST は代表的な Web アーキテクチャスタイルの一つです。

- Design Issues: Linked Data

Linked Data の原則などが記述されている文書です。

- W3C SWEO Community Project: LinkingOpenData

LOD に関するまとめサイト。ほかにもタイトルが Linking となっているものがありますが、それは歴史的経緯です。

- The Linking Open Data Cloud diagram (LOD Cloud digram)

有名な LOD の図。データセット数が多くなりすぎたとの理由で 2011 年 9 月を最後に更新が止まっています。

- Linked Open Vocabularies (LOV)

LOD Cloud で使われている語彙の統計とその可視化。

- 5 star Open Data

Linked Open Data の段階的な進め方の指標。

- Linked Data: Evolving the Web into a Global Data Space

Linked Data を体系立って学ぶことのできるテキスト。日本語訳は近代科学社より出版されています。

(1) Linked (Open) Data サイト

- Linking Open Data Cloud - the Datahub

Datahub での、LOD Cloud をつくるためのまとめグループになります。ここから手始めに探すとよいです。

●DBpedia

Wikipedia 英語版から構造化データを抽出して LOD にしたサイト。

●DBpedia Japanese

Wikipedia 日本語版から構造化データを抽出して LOD にしたサイト。ほかに 10 か国語以上のチャプタがコミュニティによって運営されています。

●日本語 Wikipedia オントロジー

Wikipedia 日本語版などから汎用的なオントロジーを構築して LOD として公開しています。同様に英語版には YAGO があります。

●LinkedGeoData

OpenStreetMap のノードと道データを主に LOD にしたサイトです。

●GeoNames

世界の地名のデータの LOD です。

●Linked data - Data.gov.uk

教育, 法令, 地理など, さまざまな LOD を提供しています。

●British Museum Semantic Web Collection Online

大英博物館のコレクション。

●Linked Open Data - data.europeana.eu

ヨーロッパ中の文化芸術施設のデータを集めている Europeana プロジェクトの LOD 試験版

●Web NDL Authorities

国立国会図書館典拠データ検索・提供サービス

●Linked Life Data

生命科学関連の主要な LOD をまとめたサイト

●The New York Times Linked Open Data

New York Times がニュースで用いている語彙の Linked Open Data

(2) SPARQL

LOD を実際に利用する方法としてよく使われているのが SPARQL という問合せ言語です。SPARQL は標準化されたデータアクセス用 Web API と捉えることができます。問合せを受け付ける場所を SPARQL エンドポイントと呼んでおり、LOD サイトの多くが SPARQL エンドポイントも提供しています。

●api/sparql - the Datahub

Datahub にて api/sparql で登録されている SPARQL エンドポイントのリスト。

●SPARQL Endpoints Status

上記の api/sparql で登録されている SPARQL エンドポイントの監視サイト

●SPARQL 1.1 Overview

W3C Recommendation, 21 March 2013, 全部で 11 個ある SPARQL 1.1 の W3C 文書のまとめになります。

●SPARQL CURRENT STATUS

W3C による SPARQL の現状についてのまとめです。

●Learning SPARQL

O'Reilly Media 社から出版されている SPARQL 解説本のサイトです。オンラインに無料で SPARQL のサンプルコードが豊富にありますので、それを見るだけでも参考になります。

●the SPARQL FAQ

SPARQL を策定している W3C RDF Data Access WG (DAWG) の Chair である Lee Feigenbaum による FAQ です。

●SPARQL by Example

同じく Lee Feigenbaum による SPARQL のチュートリアルです。

●RDF クエリ言語 SPARQL

神崎正英さんによる SPARQL が W3C 勧告になるまでの道のりのお話です。

●YASGUI

Web ベースの SPARQL クライアント。どの SPARQL エンドポイントにも問い合わせることができるので試すのに便利です。

(3) 関連仕様

- Hypertext Transfer Protocol—HTTP/1.1
IETF RFC 2616, Jun. 1999
- Uniform Resource Identifier (URI) : Generic Syntax
IETF RFC 3986, Jan. 2005
- Internationalized Resource Identifiers (IRIs)
IETF RFC 3987, Jan. 2005
- Cool URIs for Semantic Web
W3C Interest Group Note, 03 Dec. 2008
- RDF Primer
W3C Recommendation, 10 Feb. 2004, Resource Description Framework (RDF) の入門書です。RDF の文書はこれを含めて六つあります。
- Turtle: Terse RDF Triple Language
W3C Candidate Recommendation, 19 Feb. 2013, SPARQL と互換性のある文法で書ける RDF シリアライゼーションです。
- JSON-LD 1.0
W3C Last Call Working Draft, 11 April 2013, JSON 形式での RDF シリアライゼーションの一つです。
- Linked Data Platform
W3C Working Draft, 7 March 2013
- DCMI Metadata Terms
DCMI Recommendation, 14 Jun. 2012, よく使われる語彙の一つ。
- SKOS Simple Knowledge Organization System Reference
W3C Recommendation, 18 Aug. 2009, よく使われる語彙の一つ。
- Schema.RDF.org
主な検索エンジン会社が利用する語彙である Schema.org を Linked Data で扱うためのサイト。

4. ツールライブラリ

Linked Open Data に関連するツールは、Semantic Web Development Tools や LinkedData.jp: ツールにまとめられています。ここではライブラリ、解析・可視化、データ作成・公開支援、RDF Store について代表的なものを取り上げることにします。

(1) ライブラリ

RDF や SPARQL はさまざまな言語でライブラリとしてサポートされています。代表的なものを以下にあげます。

- Redland
C 用ライブラリ。Perl/PHP/Python/Ruby の binding あり。
- Ruby RDF Project
Ruby 用ライブラリ群
- Perl and RDF
Perl 用ライブラリ群
- SPARQL Endpoint interface to Python
Python の SPARQL クライアント実装
上記以外のライブラリは以下からたどれます。
- W3C Wiki: Category:Programming Language
- LinkedData.jp Wiki: ライブラリ

(2) 解析・可視化

- Sgvizler
JavaScript 用の SPARQL を使った可視化ライブラリ。
- Gephi SemanticWebImport
グラフ可視化・解析ツールである Gephi の RDF・SPARQL 用プラグイン。

- Semscape

グラフ可視化・解析ツールである Cytoscape の SPARQL 用プラグイン。

- R SPARQL

統計・可視化ソフトウェアである R 用の SPARQL プラグイン。

4.3 データ作成・公開支援

- RDF Refine

データクレンジングツールである OpenRefine の RDF 拡張。SPARQL による Reconciliation や RDF 出力などができる。

- D2RQ

RDB のデータを Linked Data として公開するためのツール。

- Pubby

SPARQL エンドポイントのデータを Linked Data として公開するためのツール。

- Link Data

RDF 作成とそのアプリケーション作成プラットフォーム。

- SparqlEPCU

RDF を作成・登録して SPARQL エンドポイントを作成できるサイト。

4.4 RDF Store

RDF を格納して処理できるデータベースです。Triple Store, Quad Store, RDF Database などいろいろな呼び方があります。代表的な実装は以下のとおりです。

- OpenLink Software Virtuoso Universal Server

- Ontotext OWLIM

- AllegroGraph

- OpenRDF Sesame

- 4store

- Apache Jena TDB

- BigData

- Mulgara

一覧は以下から辿ることができます。

- SparqlImplementation #QueryEngine

- Wikipedia: Triplestore

5. 会議・コミュニティ

Linked Open Data に関連する国内外の会議やコミュニティなどをあげます。国際会議のように毎年 Web サイトが変わるものは 2013 年 7 月時点で最新のサイトへのリンクとしました。

- Journal of Web Semantics

- Semantic Web Journal (SWJ)

- International Journal on Semantic Web and Information Systems (IJSWIS)

- International Semantic Web Conference (ISWC)

- Extended Semantic Web Conference (ESWC)

- I-SEMANTICS

- International World Wide Web Conference (WWW)

- Joint International Semantic Technology Conference (JIST)

- International Semantic Web Applications and Tools for Life Sciences Workshop (SWAT4LS)

- ACM Web Science Conference (WebSci)

- WikiCFP: Linked Data

- Linked Data - Connect Distributed Data across the Web

- 人工知能学会 セマンティックウェブとオントロジー研究会

- セマンティック Web 委員会

- Linked Open Data チャレンジ Japan
- LinkedData 勉強会

関連団体

- Open Knowledge Foundation (OKFN)
Open Data を推進するイギリスの非営利団体
- Open Knowledge Foundation Japan (OKFJ)
Open Knowledge Foundation の日本支部
- Open Data Institute
Open Data に関する研究やインキュベーションなどを促進する団体
- World Wide Web Foundation (W3F)
Web 普及のための財団
- World Wide Web Consortium (W3C)
Web に関連する仕様策定をする団体
- The Internet Engineering Task Force (IETF)
インターネット技術の仕様策定をする団体
- Linked Open Data Initiative
日本での LOD の普及促進を目的としている非営利団体

6. おわりに

本稿では Linked Open Data に関連する情報を紹介しました。Linked Open Data は技術的には World Wide Web の仕組みを踏襲したものですが、そのうえで人工知能、メディア、文化芸術、生命科学、行政、地理空間などさまざまな分野を結びつける学際的な領域として発展しています。このブックマークが Linked Open Data の世界を理解する一助になれば幸いです。