



第3回

Linked Open Data(LOD) 活用ワークショップ

RDFデータベース(トリプルストア)の導入

古崎 晃司 (大阪大学)

資料置き場 <http://bit.ly/2Pxg36q>



ハンズオンの概要

ねらい

- RDFデータベース(トリプルストア)を導入し, SPARQL クエリを実行できることを確認する.
- 導入するトリプルストア
 1. **Fuseki**: Javaベースのトリプルストア
 2. **DyDra** (<https://dydra.com>) : オンラインサービス

ハンズオンの内容

1. 格納するRDFデータの準備
2. Fusekiの利用
3. DyDraの利用(参考)



1. 格納するRDFデータの準備

ナレッジグラフ推論チャレンジ × Kozaki - X

① 保護されていない通信 | challenge.knowledge-graph.jp/index.html



1. 格納するRDFデータの準備

Screenshot of a GitHub repository page for 'Challenge/rdf/SpecledBand'.

The repository details:

- Owner: KnowledgeGraphJapan / Challenge
- Branch: master
- Issues: 0
- Pull requests: 0
- Projects: 0
- Wiki
- Insights
- Settings
- Unwatched (7)
- Starred (5)
- Forked (2)

The commit history shows the following files:

- ZIPファイルをアップ ZIPファイルをアップ 5 minutes ago
- 文章の抽出箇所 + 要約を追加 28 days ago
- SpecledBandのナレッジグラフ公開 3 months ago
- SpecledBand.nt 3 months ago
- SpecledBand.rdf 3 months ago
- SpecledBand.tsv 3 months ago
- SpecledBand.ttl 3 months ago
- SpecledBand.xlsx 3 months ago
- kgc.owl 3 months ago
- readme.md 28 days ago
- sparql.html 3 months ago

A red box highlights the file 'Specled Band.zip'. A yellow callout box points to it with the text:

「Specled Band.zip」
をダウンロード
→ZIPファイルを解凍

At the bottom of the screenshot, there is another text box:

「まだらの紐」(SpecledBand) のナレッジグラフ



Fusekiのインストール

1. <https://jena.apache.org/download/> から
Apache Jena Fusekiをダウンロード
2. Qiitaのサイト(by山本さん)を参考にして
Fusekiをインストール(※)
 - Windows編
 - <http://tinyurl.com/fuseki-win10>
 - Mac OS編
 - <http://tinyurl.com/fuseki-MacOSX>

※Javaがインストールされている必要があります.



Fusekiへのデータセットの作成①

The screenshot shows two browser windows. The main window is titled "Apache Jena Fuseki" and has the URL "localhost:3030/index.html". It displays the message "Datasets on this server" and "There are no datasets on this server yet. Add one." Below this, there is a sidebar with links: "Dataset" (Run query), "Manage datasets" (selected), and "Help" (Summary). A large blue arrow points from the "manage datasets" link in the top bar of the main window down to the "add new dataset" button in the "Manage datasets" section of the second window.

The second window is also titled "Apache Jena Fuseki" and has the URL "localhost:3030/manage.html?tab=datasets". It is titled "Manage datasets" and contains the instruction "Perform management actions on existing datasets, including backup, or add a new dataset." It lists "existing datasets" and features a prominent red box around the "add new dataset" button. Below this, there is a table with a single row for "/KGC", which includes "remove", "backup", and "upload data" buttons.



Fusekiへのデータセットの作成②

Apache Jena Fuseki - mar

localhost:3030/manage.html?tab=new-dataset

Apache Jena Fuseki

dataset manage datasets help

Manage datasets

Perform management actions on existing datasets, including backup, or add

+ existing datasets add new dataset

Dataset name: KGC

Dataset type: In-memory – dataset will be recreated when Fuseki restarts, but contents will be lost
 Persistent – dataset will persist across Fuseki restarts

create dataset

データセット名の入力
(例: KGC)
※名前は任意

アップロードしたデータを、Fuseki
の再起動後も使用したい場合は
「Persistent」を選択



Fusekiへのデータのアップロード

The screenshot shows two browser windows. The left window is a smaller view of the same interface, focusing on the 'upload files' button which is highlighted with a red box. The right window is a larger view of the 'dataset.html?tab=upload&ds=/KGC' page. It shows the 'Dataset' dropdown set to '/KGC'. Below it, the 'Upload files' section has a 'Destination graph name' input field ('Leave blank for default graph') and a 'Files to upload' section. A green file 'SpeckledBand.ttl' (313.9kb) is selected. The status message below says 'Result: success. 4381 triples'. A large blue arrow points from the 'upload files' button in the smaller window up towards the 'upload now' button in the larger window. A red box highlights the 'select files...' button in the 'Files to upload' section of the larger window. A blue box at the bottom contains Japanese text instructions.

Apache Jena Fuseki - insp

localhost:3030/dataset.html?tab=upload&ds=/KGC

Apache Jena Fuseki

dataset manage datasets

Dataset: /KGC

query upload files edit info

Upload files

Load data into the default graph of the currently selected dataset, or the given named graph. You may upload any RDF format, such as Turtle, RDF/XML or TRIG.

Destination graph name Leave blank for default graph

Files to upload + select files... upload all

SpeckledBand.ttl 313.9kb
Result: success. 4381 triples

+ select files... upload all

upload now remove

ファイルを選択してUpload
※今回は「SpeckledBand.ttl」を使用
(SpeckledBand.nt, SpeckledBand.rdfも可)



FusekiのSPARQLエンドポイント

Apache Jena Fuseki - insp Koza

localhost:3030/dataset.html?tab=upload&ds=/KGC

Apache Jena Fuseki dataset manage data

Dataset: /KGC

Available services

File Upload: <http://localhost:3030/KGC/upload>
Graph Store Protocol: <http://localhost:3030/KGC/data>
Graph Store Protocol (Read): <http://localhost:3030/KGC/get>
HTTP Quads: <http://localhost:3030/KGC/>

SPARQL Query: <http://localhost:3030/KGC/query>
SPARQL Query: <http://localhost:3030/KGC/sparql>
SPARQL Update: <http://localhost:3030/KGC/update>

Statistics

Name	Overall good	Overall bad	SPARQL Update

データセットの情報の確認

SPARQLエンドポイントのURL



FusekiのSPARQLエンドポイント

WebブラウザからSPARQLクエリで検索する場合

query

EXAMPLE QUERIES
Selection of triples Selection of classes

PREFIXES
rdf rdfs owl xsd +

SPARQL ENDPOINT
<http://localhost:3030/KGC/query>

```
1 #死んだ人が出てくるシーン一覧
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
4 PREFIX kd: <http://kgc.knowledge-graph.jp/data/SpecifiedBand/>
5 SELECT ?id ?s
6 WHERE [
7     ?id kgc:hasPredicate kd:die ;
8     kgc:subject ?s .
```

ナレッジグラフ推論チャレンジのデータセット公開ページにあるサンプルクエリを試してみる
<http://challenge.knowledge-graph.jp/sparql.html>

クエリ実行ボタン



他のツールからの利用

Linked Data Visualization x

← → ⌂ 保護されていない通信 | knowledge

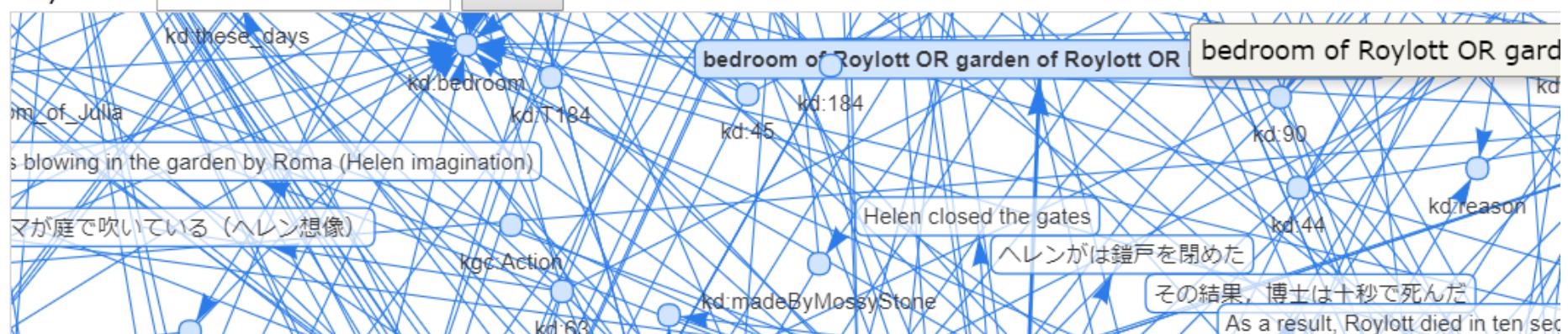
ナレッジグラフ推論チャレンジ
の可視化ツールからの利用
<http://knowledge-graph.jp/visualization/>

Linked Data visualization tool

```
#必ず?s ?p ?oの結果が帰ってくるようなクエリを書いてください
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX kgc: <http://kgc.knowledge-graph.jp/ontology/kgc.owl#>
PREFIX kd: <http://kgc.knowledge-graph.jp/data/SpeckledBank/>
SELECT ?s ?p ?o
WHERE {
  ?s ?p ?o.
```

SPARQL Endpoint: load stop clear

Keyword:



FusekiのSPARQLエンドポイントの
URLを入れて「load」を試す



参考:Fusekiのデータ更新

- 第1回LOD活用ワークショップの山本さんの資料
「トリプルストアの導入」
<https://github.com/KnowledgeGraphJapan/LODws1st/tree/master/RDF-DB>
 - Qiitaの記事(by山本さん)
<http://tinyurl.com/fuseki-intro>
- を参考にして下さい。



DyDraの利用

1: アカウント作成

開催趣旨

- LODを利用する技術を学びたいと考えている方を対象に、ハンズオン形式による実践的な技術習得の場を提供する。

プログラム概要

- 13:30-15:30【第1部】SPARQL入門ハンズオン
 - LODの基礎/LOD用検索用クエリ言語SPARQLの基礎
古崎晃司(大阪大学)
- 16:00-18:00【第2部】LOD活用ツールハンズオン
 - RDFデータベース(トリプルストア)の導入
古崎晃司(大阪大学)
 - LOD活用ライブラリの紹介
加藤文彦(国立情報学研究所)
 - OpenRefineによるWikidataの編集
山本泰智(ライフサイエンス統合データベースセンター)



DyDraの利用

1: アカウント作成

開催趣旨

- LODを利用する技術を学びたいと考えている方を対象に、ハンズオン形式による実践的な技術習得の場を提供する。

プログラム概要

- 13:30-15:30【第1部】SPARQL入門ハンズオン
 - LODの基礎/LOD用検索用クエリ言語SPARQLの基礎
古崎晃司(大阪大学)
- 16:00-18:00【第2部】LOD活用ツールハンズオン
 - RDFデータベース(トリプルストア)の導入
古崎晃司(大阪大学)
 - LOD活用ライブラリの紹介
加藤文彦(国立情報学研究所)
 - OpenRefineによるWikidataの編集
山本泰智(ライフサイエンス統合データベースセンター)



参考:DyDraの利用

- DyDra (<https://dydra.com>)

- RDFデータベース(トリプルストア)を提供するオンラインサービス
- 無料で**100万トリプル**のデータを利用可能。
(クエリは毎秒5回まで)
- アカウント登録必要
(少し時間がかかる?)
- 小規模なRDFデータの利用に便利

