LODの作り方・使い方

松村 冬子 fuyuko@linkedopendata.jp @wintermelon

June 26^{th,} 2014 リンクト・オープン・データ・イニシアティブ 青山学院大学





おさらい:RDFでデータ表現

- トリプル:主語・述語・目的語
- 語彙: クラス + プロパティの集合





今日の講義の流れ

- LODを作る
 - -表からLODを作る
 - 外部のLODとつなぐ
 - ー作ったLODを公開する
- LODを使う
 - -SPARQLでほしいデータを取得する



今日の材料

相模原市の施設データ(CSV形式)

相模原市 Sagamihara City RAED 8//245			文字サイズ 編 Google [™] カスタム検す	諸小 元に戻す 拡大 背: 素	景色の変更 A A A A Q 検索 >検索の使い方
トップページ	4 暮らしの情報	🗐 市政情報	🔛 観光・文化	産業・ビジネス	♀ 施設マップ

現在の位置 : <u>トップページ</u>、<u>市政情報</u>、相模原市オープンデータライブラリー



http://www.city.sagamihara.kanagawa.jp/opendata/index.html



施設データ例

小・中学校などの施設データ(CSV)

http://www.city.sagamihara.kanagawa.jp/dbps_data/_material_/_files/ 000/000/028/373/04_gakkou_20140401.csv

	Α	В	С	D	E	F	G
1	施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
2	環境情報センター	〒252-0236	相模原市中央	中央区	042-769-9248	35.57026448	139.3751576
3	相模川ビレッジ若あゆ(相	〒252-0135	相模原市緑区	緑区	042-760-5445	35.57970899	139.3061113
4	城山学校給食センター	〒252-0111	相模原市緑区	緑区	042-782-3094	35.590584	139.3081984
5	青少年学習センター	〒252-0207	相模原市中央	中央区	042-751-0091	35.57467081	139.3872353
6	青少年相談センター	〒252-0239	相模原市中央	中央区	042-769-8285	35.57032874	139.3730674
7	清新学校給食センター	〒252-0217	相模原市中央	中央区	042-771-7218	35.5726655	139.357617
8	総合学習センター	〒252-0239	相模原市中央	中央区	042-756-3443(35.56969074	139.372043
9	津久井学校給食センター	〒252-0153	相模原市緑区	緑区	042-784-0522	35.5770529	139.2757365





LODを作る





3. 変換作業



1. スキーマの設計

	Α	В	С	D	E	F	G
1	施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
2	環境情報センター	〒252-0236	相模原市中央	中央区	042-769-9248	35.57026448	139.3751576
3	相模川ビレッジ若あゆ(相	〒252-0135	相模原市緑区	緑区	042-760-5445	35.57970899	139.3061113
4	城山学校給食センター	〒252-0111	相模原市緑区	緑区	042-782-3094	35.590584	139.3081984
5	青少年学習センター	〒252-0207	相模原市中央	中央区	042-751-0091	35.57467081	139.3872353
6	青少年相談センター	〒252-0239	相模原市中央	中央区	042-769-8285	35.57032874	139.3730674
7	清新学校給食センター	〒252-0217	相模原市中央	中央区	042-771-7218	35.5726655	139.357617
8	総合学習センター	〒252-0239	相模原市中央	中央区	042-756-3443(35.56969074	139.372043
9	津久井学校給食センター	〒252-0153	相模原市緑区	緑区	042-784-0522	35.5770529	139.2757365

ⓒ € 相模原市

- 何をリソースにする?
- どんなクラスが必要?
- 何をプロパティにする?



何をリソースにする?

	Α	В	С	D	E	F	G
1	施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
2	環境情報センター	〒252-0236	相模原市中央	中央区	042-769-9248	35.57026448	139.3751576
3	相模川ビレッジ若あゆ(相	〒252-0135	相模原市緑区	緑区	042-760-5445	35.57970899	139.3061113
4	城山学校給食センター	〒252-0111	相模原市緑区	緑区	042-782-3094	35.590584	139.3081984
5	青少年学習センター	〒252-0207	相模原市中央	中央区	042-751-0091	35.57467081	139.3872353
6	青少年相談センター	〒252-0239	相模原市中央	中央区	042-769-8285	35.57032874	139.3730674
7	清新学校給食センター	〒252-0217	相模原市中央	中央区	042-771-7218	35.5726655	139.357617
8	総合学習センター	〒252-0239	相模原市中央	中央区	042-756-3443(35.56969074	139.372043
9	津久井学校給食センター	〒252-0153	相模原市緑区	緑区	042-784-0522	35.5770529	139.2757365

ⓒ € 相模原市



• 行のデータ全体

•施設を表現するURIが必要



どんなクラスが必要?

- どのように語彙を選ぶか?
 - 既存の語彙をできるだけ利用(今回)

- オリジナルの語彙から策定

- 施設を表現するクラス
 - -小学校 schema:ElementarySchool
 - -中学校 schema:MiddleSchool など



クラスの分け方は 用途に合わせて

何をプロパティにする?

	A	В	С	D	E	F	G
1	施設名	郵便番号	主所	地域区分	電話番号	緯度	経度
2	環境情報センター	〒252-0236	相模原市中央	中央区	042-769-9248	35.57026448	139.3751576
3	相模川ビレッジ若あゆ(相	〒252-0135	相模原市緑区	緑区	042-760-5445	35.57970899	139.3061113
4	城山学校給食センター	〒252-0111	相模原市緑区	緑区	042-782-3094	35.590584	139.3081984
5	青少年学習センター	〒252-0207	相模原市中央	中央区	042-751-0091	35.57467081	139.3872353
6	青少年相談センター	〒252-0239	相模原市中央	中央区	042-769-8285	35.57032874	139.3730674
7	清新学校給食センター	〒252-0217	相模原市中央	中央区	042-771-7218	35.5726655	139.357617
8	総合学習センター	〒252-0239	相模原市中央	中央区	042-756-3443(35.56969074	139.372043
9	津久井学校給食センター	〒252-0153	相模原市緑区	緑区	042-784-0522	35.5770529	139.2757365

ⓒ € 相模原市





プロパティと既存語彙のマッピング

施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
中央中学校	〒252-0236	相模原市中央	中央区	042-755-0071	35.56947172	139.3773524

- 施設名 rdfs:label
- 郵便番号 schema:postalCode
- 住所 schema:address / 前回資料を参考
- 地域区分 ?????
- 電話番号 schema:telephone
- 緯度 geo:lat
- 経度 geo:long



&

似たLODの

語彙を真似る

プロパティと既存語彙のマッピング

施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
中央中学校	〒252-0236	相模原市中央	中央区	042-755-0071	35.56947172	139.3773524

- 施設名 rdfs:label
- 郵便番号 schema:postalCode
- 住所 schema:address
- 地域区分 ?????
- 電話番号 schema:telephone
- 緯度 geo:lat
- 経度 geo:long



適当な語彙

がない

プロパティと既存語彙のマッピング

施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
中央中学校	〒252-0236	相模原市中央	中央区	042-755-0071	35.56947172	139.3773524

• 施設名 rdfs:label

※メジャーなPrefixは <u>prefix.cc</u>で調べられる

- 郵便番号 schema:postalCode
- 住所 schema:address
- 地域区分 sgmhr:areaCategory
- 電話番号
- schema:telephone
- 緯度 geo:lat
- 経度 geo:long









各リソースを表すURIを決める

一意にリソースを指定できる ユニークな値を利用するとよい

例)

- リソースの名前を使う http://sgmhr.jp/resource/Chuo_Chugakko
- リソースを示すIDを使う http://sgmhr.jp/place/1234



このスキーマにしたがって表現すると

施設名	郵便番号	住所	地域区分	電話番号	緯度	経度
中央中学校	〒252-0236	相模原市中央	中央区	042-755-0071	35.56947172	139.3773524







表データからRDFへの変換

- 表データ = xls, csv, tsv など
- 変換方法
 - -変換ツールを使う
 - LinkData
 - OpenRefine + RDFRefine
 - スクリプトを書く
 - お好みの言語でどうぞ
 - 余談: RDBからの変換ツールもある
 - R2RML など



LinkData http://linkdata.org/



- 国内のオープンデータ活動支援プラットフォーム
- 他のデータを流用したデータやアプリを作れる
- 変換手順(チュートリアル参照)
 - 1. ひながたファイルの作成
 - 各列のプロパティやデータ型を選べる
 - xlsとしてダウンロード or Google Driveに作成
 - 2. ひながたファイルにデータを入力
 - 3. データのアップロード・公開設定
 - 4. Turtle (RDF) がダウンロード可能になる



OpenRefine + **RDF Refine**

http://openrefine.org/

http://refine.deri.ie/

• OpenRefine (|⊟: Google Refine)

- データの整形やクリーニングを行うツール

• RDF Refine



- OpenRefineの拡張機能
- 列とプロパティの語彙を対応づけRDFに変換
- DBpediaなどに含まれる同一のリソースを 探してつなぐことが可能



OpenRefineの導入

- Windows編(Mac版, Linux版もあり)
 - ダウンロードして解凍し、適当な場所におく <u>http://openrefine.org/download.html</u>(今回はVer2.5)
 - 解凍した中の google-refine.exe を実行
 - – 自動的にブラウザで
 Open Refineが起動

Google refine	A power tool for working with messy data.					
Create Project Open Project Import Project	Create a project by importing data. What kinds of data files can I import? TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.					
	Get data from	Locate one or more files on your computer to upload:				
	This Computer	ファイル選択 選択されていません				
	Web Addresses (URLs)	Next »				
	Clipboard					
	Google Data					
Version 2.5 [r2407]						
Help About						



RDF Refineの導入

- Open Refine 起動後に行う
 - ダウンロードして解凍 <u>http://refine.deri.ie/</u>(今回はVer0.8.0)
 - 解凍した中の rdf-extention フォルダを OpenRefineのインストールフォルダの webappフォルダの下のextentionsフォルダの下 にコピーしOpenRefineを再起動





表データの読込方法

例)相模原市小・中学校など所在地一覧

CSVファイルURL

http://www.city.sagamihara.kanagawa.jp/dbps_data/_material_/_files/000/000/028/373/04_gakkou_20140401.csv

- 方法1:CSVファイルをアップロード
- 方法2:Web上の表データのURLを指定
- 方法3:クリップボードでコピー&ペースト
- 方法4:Google Spreadsheetのデータを流用



表データの読込(方法1:ファイルから)

Google refine A power tool for working with messy data. Create a project by importing data. What kinds of data files can I import? Create Project TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data docur Open Project formats can be added with Google Refine extensions. Import Project Get data from Locate one or more files on your computer to upload: 方法1→ $(\mathbf{2})$ This Computer (3) ファイル選択 04 gakkou 20140401.csv 方法2→ Web Addresses (URLs) (**4**) Next » PCに保存されている 方法3→ Clipboard 表データの Google Data 方法4→ ファイルを選択



表データの読込(方法2:Webから)





表データの読込確認・文字コード編集

⑤ プロジェクト名 ⑥次へ Google refine A power tool for working with messy data. « Start Over Configure Parsing Options Project name 04 gakkou 20140401 csv Create Project » Create Project Open Project 施設名 郵便番号 住所 地域区分 電話番号 緯度 経度 Import Project 環境情報センター 相模原市 〒252-1 見1-3-0236 く読み込めているか確認 〒252-2. 相模川ビレッジ若あゆ(相模川自然 の村野外体験教室) 0135 文字化け → ② 城山学校給食センター 3 ₹252-相模》 0111 1673-1 列の区切り (3) \rightarrow 相模原词 4 青少年学習センター **T252-**新町3-0207 青少年相談センター 5. **T252-**相模原 で再確認 (3)修正 (4) 0239 3-13-1 清新学校給食センター 相模原词 **T252-**6. 0217 i涌2-2-1 211 A 332 THE (4 Parse data as Character encoding Update Preview CSV / TSV / separator-based files Columns are separated by Ianore 0 line(s) first at beginning of commas (CSV) Line-based text files file tabs (TSV) Fixed-width field text files Parse line(s) custom , PC-Axis text files next as column Escape special characters with \ headers ISON files



プロジェクト作成完了

Google refine 相模原市小・中学校など Permalink

Extensions: Freebase RDF -	159 rows	Facet / Filter Undo / Redo o
0 25 50 rows Edit RDF Skeleton	Show as: rows records Show: 5 10 25 50 row	
郵便番号 🔽 住所 Reset RDF Skeleton	▼ All ▼ 施設名 ▼ 郵便番号	Using facets and filters
252-0236 相模原市 Add reconciliation service	☆ 🧠 1 環境情報センター 〒252-0236 1	
中央区留 士見1-3- 41		Use facets and filters to select subsets of your data to act on. Choose facet and
252-0135 相模原市 録区 042-760-5445 35.5797089894 録区大島 3497-1	✓ □ 2. 相模川ビレッジ若あゆ(相模 川自然の村野外体験教室) 〒252-0135 単	of each data column.
252-0111 相模原式 録区 042-782-3094 35.5905839987 録区川県 1673-1	- 3. 城山学校給食センター 〒252-0111 4	表の各列の操作ボタン
252-0207 相模原方 中央区 042-751-0091 35.5746708097 中央区天 部新町 3- 15	→ 4. 青少年学習センター 〒252-0207 月 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日	(列の追加・削除, 列名の変更などが可能)
1-7人区留 士見1-3- 41 1 252-0135 相模原市 绿区大島 3497-1 緑区 042-760-5445 35.5797089 252-0111 相模原市 绿区川月 1673-1 緑区 042-782-3094 35.5905839 252-0207 相模原市 中央区天 部新町3- 15 中央区 042-751-0091 35.5746708	□ 2. 相模川ビレッジ若あゆ(相模 川自然の村野外体験教室) 〒252-0135 □ 3. 城山学校給食センター 〒252-0111 □ 4. 青少年学習センター 〒252-0207	Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

列とプロパティのマッピング
RDFボタンから RDF Skelton を編集



RDF Refineのメニューボタン

インストールが失敗)

(なければRDF Refineの

列とプロパティのマッピング

RDF Schema Alignment



RDFに変換(RDF/XML)

Open Excort	<pre> </pre>
Export project	xmlns.geo- http://www.ws.org/2003/01/geo/wgso4_pos#xmlns.roar- http://xmlns.com/roar/0.1/ 3 xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
	▼ <rdf:description_rdf:about="http: 0"="" place="" sgmhr.jp=""></rdf:description_rdf:about="http:>
Tab-separated value	<pre><rui:type <rufs:label="" rui:tesource="http://schema.org/Educationatorganization">環境情報センター</rui:type></pre>
Comma-separated value	<pre><schema:postalcode>=252-0236</schema:postalcode></pre>
HTML table	<pre><schema:address>相模原市中央区富士見1-3-41</schema:address></pre>
TTIME table	<sgmhr:areauategory>中央区</sgmhr:areauategory>
Excel	<pre> Schema:terephone/04z=709=9z4835.5702644773 </pre>
ODE enreadeheat	<pre><ge0:tat tuf.uatatype="http://www.wo.org/2001/XMLSchema#double">00.07020447700/ge0.Tat/ <ge0:long_rdf:datatype="http: 2001="" www.w3.org="" xmlschema#double"="">139_375157631L</ge0:long_rdf:datatype="http:></ge0:tat></pre>
ODF spreadsneet	<pre></pre>
Triple loader	<pre>rdf:Description rdf:about="http://sgmhr.jp/place/1"></pre>
inpic loader	<pre><rdf:type rdf:resource="http://schema.org/EducationalOrganization"></rdf:type></pre>
MQLWrite	<rdts:label>相模川ビレッジ右あゆ(相模川目然の村野外体験教室)</rdts:label>
	<pre><schema:postalloade <schema:address="" schema:postalloade="" 〒202-0130<=""></schema:postalloade></pre>
Custom tabular exporter	<pre></pre>
Templating	<pre><schema:telephone>042-760-5445</schema:telephone></pre>
	<pre><geo:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#double">35.5797089894</geo:lat></pre>
RDF as RDF/XML	<pre><geo:long_rdf:datatype="http: 2001="" www.w3.org="" xmlschema#double"="">139.306111336</geo:long_rdf:datatype="http:></pre>
DDE ee Turtle	
RDF as future	<pre><rul:description <="" pre="" rul:about="http://sgmmi.jp/prace/2"></rul:description></pre>
	<pre></pre>
	<schema:postalcode>=252-0111</schema:postalcode>
	<schema:address>相模原市緑区川尻1673-1</schema:address>
	<sgmhr:areacategory>緑区</sgmhr:areacategory>
	<pre><schema:telephone>U42-782-3094</schema:telephone> </pre>
	seetlong_rdf:datatype="http://www.w3.org/2001/XMLSchema#double">30.090083998/\/geotlong> <geotlong_rdf:datatype="http: 2001="" www.w3.org="" xmlschema#double"="">130_308108/13</geotlong_rdf:datatype="http:>
	<pre></pre>
	<pre>rdf:Description rdf:about="http://sgmhr.jp/place/3"></pre>
	<pre><rdf:type rdf:resource="http://schema.org/EducationalOrganization"></rdf:type></pre>



RDFに変換(Turtle)

Open Export	@prefix sgmhr: <http: sgmhr.jp=""></http:> .	I
Export project	@prefix schema: <http: schema.org=""></http:> . @prefix rdfs: <http: 01="" 2000="" rdf-schema#="" www.w3.org=""> .</http:>	
Tab-separated value Comma-separated value HTML table Excel ODF spreadsheet Triple loader MQLWrite Custom tabular exporter	<pre>@prefix rdfs: <http: 01="" 2000="" rdf-schema#="" www.w3.org=""> . @prefix geo: <http: 01="" 2003="" geo="" wgs84_pos#="" www.w3.org=""> . @prefix foaf: <http: 0.1="" foaf="" wmlns.com=""></http:> . @prefix xsd: <http: 2001="" www.w3.org="" xmlschema#=""> . @prefix owl: <http: 07="" 2002="" owl#="" www.w3.org=""> . @prefix rdf: <http: 02="" 1999="" 22-rdf-syntax-ns#="" www.w3.org=""> .</http:></http:></http:></http:></http:></pre>	
Templating	geo:lat "35.5702644773"^^xsd:double ;	
RDF as RDF/XML	geo:long ~139.3/515/631~~~xsd:double .	L
RDF as Turtle	<pre><http: 1="" place="" sgmhr.jp=""> a schema:EducationalOrganization ; rdfs:label "相模川ビレッジ若あゆ(相模川自然の村野外体験教室)"; schema:postalCode "〒252-0135"; schema:address "相模原市緑区大島3497-1"; sgmhr:areaCategory "緑区"; schema:telephone "042-760-5445"; geo:lat "35.5797089894"^^xsd:double; geo:long "139.306111336"^^xsd:double.</http:></pre>	
	<pre><http: 2="" place="" sgmhr.jp=""> a schema:EducationalOrganization ; rdfs:label ~城山学校給食センター~; schema:postalCode ~〒252-0111~; schema:address ~相模原市緑区川尻1673-1~; sgmhr:areaCategory ~緑区~; schema:telephone ~042-782-3094~; geo:lat ~35.5905839987~^xsd:double ; }</http:></pre>	ative

外部のデータとつなぐには

- Linked Open Dataにしたい
- つながりの種類は色々
 - owl:sameAs (同じ事物・概念を指す)
 - dcterms:references (ある事物・概念を参照する)
 - skos:related (ある事物・概念に関連する)
- つながりの見つけ方
 - オントロジーマッチングやインスタンスマッチング
 などの研究
 - 今回は単純に字面の類似性によって探す



Reconciliation機能でつなぐ

- 目的:他のLODに含まれる同じ事物を指す
 リソースとマッチング
- 基準: ラベルやプロパティの値の類似性
- 例) DBpedia Japaneseに含まれる同じ施設の
 データをつなぐ
- 今回はSPARQLエンドポイントを利用



Reconciliationの設定

• Reconciliation先のSPARQLエンドポイントを設定



一致する事物のデータを探す

例)同じ施設名がrdfs:labelに含まれる
 DBpedia Japaneseのデータ

▼ 施設名		• \$	Reconcile column "施設名	, "		
Facet		〒252	Freebase Query-based	Reconcile eac		
Edit cells	_ 奥川自然の村野外体験教室〉	〒252 〒252	Sindice	owl:Thin	回しナータを検察開始 検索性2000	
Edit column		∓ 252	Freebase Reconciliation	dbo:Arcl bttp://db	快糸仏儿が回回上即に衣小	
Transpose 🕨	_	∓252 ∓252	DBpedia Japanese 🗵	 schema: http://sci 	Reconcile cells in column 施設名 to type	
\$2	3	1202	DBpedia	dbo:Buil http://db	http://www.w3.org/2002/07/owl#Thing	
View		12		 dbo:Plat http://db dbo:Wikida 		
Reconcile	Start reconciling.	this		5 type	eを指定して検索しない	
DBpe	gainst type:					
(右側画面が表示されるまで ▲ Auto-match candidates with high confidence						
一少	し時間がかれ	うりる	ます)		(6)	
			Add Standard Service	amespaced Service	Start Reconciling	
Reconciliation完了





候補にマッチするものがある場合







マッチしたURIを格納する列を作成



マッチしたURIを格納&RDF生成







作ったLODを公開する

- RDFダンプ
- **RDF**ストア
 - Virtuoso
 - -OWLIM
 - -Allegro Graph
 - -ARC2
 - SparqIEPCU
- 参照解決可能



SparqIEPCU <u>http://lodcu.cs.chubu.ac.jp/SparqIEPCU/</u>

・RDFの格納・検索が可能

SparqlEPCU		
A Home A Tutorial -	🖆 Project 🔲 Schema 🖾 Demo 🖾 Contact Fuyuko Ito Matsumura 🛪	
Login ようこそFuyuko Ito Matsumuraさん ログアウト	はじめに このサイトはLinked Open Data(LOD)作成及び活用の支援サイトです。	
RDFmgr Download RDFmgr	 、 LODアプリを体験したい LODアプリを体験したい LODアプリサイトをJavaScriptで簡単に作りたい AのためにSecondEPCUではい下を提供します 	
<script type="text/java Download JS sample code.	 そのためにSparqiEPCOでは以下を提供します 1. インターネットで共有するRDF格納サイト 2. 登録したRDFデータのSparql検索を行うWebページ 3. RDF登録・検索のためのJavaScriptライブラリ(RDFmgr) 4. RDF登録・検索するJavaScriptサンプルコード 5. LODアプリサイトのサンプル 	
お知らせ ・ Twitter EP SparglEPCU	現在Internet Explorerでは正常に動作しない場合があります。	n Data Initiati



SparqIEPCUにデータを登録

	・ プロ
	画
	登
TURTLE	
@prefix owl: <http: 07="" 2002="" owl#="" www.w3.org=""> . @prefix rdf: <http: 02="" 1999="" 22-rdf-syntax-ns#="" www.w3.org=""> .</http:></http:>	• _ t
<http: 0="" place="" sgmhr.jp=""> a schema:<u>EducationalOrganization</u> ; rdfs:label "環境情報センター" ; schema:postalCode "〒252-0236" ;</http:>	_
schema:address "相模原市中央区富士見1-3-41"; sgmhr:areaCatergory "中央区"; schema:telephone "042-769-9248";),
 ファイルによる登録 ファイルによる登録 ファイルによる登録 ファイルによる登録 	• N3
<u>Choose File</u> No file Chosen 拡張子(CSV→.csv, N3→.n3, RDF/XML→.xml, TURTLE→.ttl, N-TRIPLE→.nt)にご注意ください。	Tur
④ 保存してRDFを生成	

プロジェクトの 画面下部の「データ 登録」タブを選択

ピー&ペースト or アイルでの登録

• N3, RDF/XML, Turtle, N-tripleなど



SparqIEPCUでのデータ登録確認

SPARQL検索 UPDATE デ SPARQL検索 ① SPA	^{ータ登録} ダウンロード ブロ ARQL検索タブ	^{ジェクト設定} ブを選択	
select * where {?s ?p ?o} LIMIT 10 検索実行 検索結果をクリア	∞ ② とりあえ	ず検索実行	
S	p	0	
http://sgmhr.jp/place/12	rdf:type	http://schema.org/EducationalOrgan ization	
http://sgmhr.jp/place/12	rdfs:label	青根小学校	
http://sgmhr.jp/place/12	http://schema.org/postalCode	〒252-0162	プルが
http://sgmhr.jp/place/12	http://schema.org/address	相模原市緑区青樹、表示され	っていわげ
http://sgmhr.jp/place/12	http://sgmhr.jp/areaCatergory		
http://sgmhr.jp/place/12	http://schema.org/telephone	042-787-2 公開	成切!
http://sgmhr.jp/place/12	geo:lat	35.54375	
http://sgmhr.jp/place/12	geo:long	139.129223	



LODを使う



LODを使うための技術

- RDFストア
 RDFを格納するデータベース
- SPARQL
 RDFの中を検索するための言語
- SPARQLエンドポイント RDFストアのデータ入出力インタフェース



d Open Data Initiative

LODとRDBの比較





SPARQLの基礎知識



SPARQL

SPARQL Protocol And RDF Query Language

Recommendation

N3C

- SQLに似てる
- W3Cで標準化
 - -2008年1月 SPARQL1.0 勧告
 - -2013年3月 SPARQL1.1 勧告



Copyright © 2013 W3C® (MIT, ERCIM, Keio, Beihang), All Rights Reserved. W3C liab





- SELECT
 取得する変数の指定
 * = 全変数
- WHERE{}
 取得したいグラフ パターンの指定 (組み合わせ可)
- ?文字列









SELECT * WHERE{ ?s ?p ?o .





SELECT * WHERE{ ?s ?p ?o .





SPARQL検索

select * where {?s ?p ?o} LIMIT 10



検索実行

検索結果をクリア

↓各変数の値の組み合わせのリストが検索結果になる

S	p	0
http://sgmhr.jp/place/12	rdf:type	http://schema.org/EducationalOrga nization
http://sgmhr.jp/place/12	rdfs:label	青根小学校
http://sgmhr.jp/place/12	http://schema.org/postalCode	〒252-0162
http://sgmhr.jp/place/12	http://schema.org/address	相模原市緑区青根1331
http://sgmhr.jp/place/12	http://sgmhr.jp/areaCatergory	緑区
http://sgmhr.jp/place/12	http://schema.org/telephone	042-787-2534
http://sgmhr.jp/place/12	geo:lat	35.54375
http://sgmhr.jp/place/12	geo:long	139.129223
http://sgmhr.jp/place/141	rdf:type	http://schema.org/EducationalOrga nization
http://sgmhr.jp/place/120	rdf:type	http://schema.org/EducationalOrga nization



グラフパターンの種類

- ・完全一致
 - 例) sgmhrp:12 rdfs:label "青根小学校".
- ・ 変数によるパターンマッチ
 - 例) sgmhr:12 rdfs:label ?label.
- 上記の組合せ
 同じ主語のトリプルが続く
 → 主語を省略して;でつなぎ一文に
 例)?s rdfs:label "青根小学校"; geo:lat ?lat; geo:long ?long.



SELECT * WHERE{ **?s** rdfs:label **?label**. } LIMIT 10



S	label
http://sgmhr.jp/place/82	鳥屋中学校
http://sgmhr.jp/place/88	根小屋小学校
http://sgmhr.jp/place/19	麻溝台中学校
http://sgmhr.jp/place/64	清新小学校
http://sgmhr.jp/place/16	青葉小学校
http://sgmhr.jp/place/58	相模台中学校
http://sgmhr.jp/place/21	新磯小学校
http://sgmhr.jp/place/0	環境情報センター
http://sgmhr.jp/place/17	旭小学校
http://sgmhr.jp/place/63	新町中学校



SELECT * WHERE{ ?s rdfs:label ?label; geo:lat ?lat; geo:long ?long. } LIMIT 10



s	label	lat	long
http://sgmhr.jp/place/82	鳥屋中学校	35.5456556953	139.217468586
http://sgmhr.jp/place/88	根小屋小学校	35.5754581182	139.273658721
http://sgmhr.jp/place/19	麻溝台中学校	35.5315216089	139.408636796
http://sgmhr.jp/place/64	清新小学校	35.5771715596	139.361787667
http://sgmhr.jp/place/16	青葉小学校	35.5544529771	139.383437418
http://sgmhr.jp/place/58	相模台中学校	35.5214278086	139.407552199
http://sgmhr.jp/place/21	新磯小学校	35.5036867935	139.383952601
http://sgmhr.jp/place/0	環境情報センター	35.5702644773	139.375157631
http://sgmhr.jp/place/17	旭小学校	35.5991	139.343408
http://sgmhr.jp/place/63	新町中学校	35.5253279407	139.437962782





lat	long
35.5544529771	139.383437418







※ SparqlEPCUではいくつかの主なPrefixが予め定義されており, rdfs や geo ではエラーが出なかった

s	label
http://sgmhr.jp/place/38	小山中学校
http://sgmhr.jp/place/71	田名小学校
http://sgmhr.jp/place/37	小山小学校
http://sgmhr.jp/place/90	光が丘小学校
http://sgmhr.jp/place/6	総合学習センター
http://sgmhr.jp/place/108	弥栄中学校
http://sgmhr.jp/place/64	清新小学校
http://sgmhr.jp/place/48	共和中学校
http://sgmhr.jp/place/4	青少年相談センター
http://sgmhr.jp/place/101	星が丘小学校

SELECT * WHERE{ ?s rdfs:label ?label; <http: areacategor<br="" sgmhr.jp="">} LIMIT 10</http:>	y> "中央区". もしくは<>でURI を囲ってもOK
s	label
http://sgmhr.jp/place/38	小山中学校
http://sgmhr.jp/place/71	田名小学校
http://sgmhr.jp/place/37	小山小学校
http://sgmhr.jp/place/90	光が丘小学校
http://sgmhr.jp/place/6	総合学習センター
http://sgmhr.jp/place/108	弥栄中学校
http://sgmhr.jp/place/64	清新小学校
http://sgmhr.jp/place/48	共和中学校
http://sgmhr.jp/place/4	青少年相談センター
http://sgmhr.jp/place/101	星が丘小学校

SPARQL 1.0



SPARQL 1.0

- データ取得
 - SELECT
 - -ASK
 - DESCRIBE
 - CONSTRUCT
- ほぼどのサーバでも使える
 - * SparqIEPCUでは対応していないものもあるので, その場合はDBpedia Japanese などでどうぞ http://ja.dbpedia.org/sparql



SELECT

 グラフパターンにマッチする 変数の値の組合せを取得

例)任意のトリプルを10件取得

```
SELECT *
WHERE{
?s ?p ?o .
}
LIMIT 10
```



ASK

グラフパターンにマッチするものが あるかどうか(YES/NO)

例) "青葉小学校"というラベルのリソースがあるか

```
ASK
WHERE{
?s rdfs:label "青葉小学校" .
}
```



DESCRIBE

- リソースに関するグラフを取得
- サーバの実装によって挙動が異なる
 - 多くの場合: リソースを主語/目的語に含むトリプルの集合
- 結果形式:RDF/XML, Turtle, ...

例) http://sgmhr.jp/place/38 に関するグラフを取得

DESCRIBE <http://sgmhr.jp/place/38>



CONSTRUCT

- 新しいグラフを作成して取得
- 結果形式:RDF/XML, Turtle, ...
- 例)ある10件のトリプルのrdfs:labelのプロパティを foaf:nameに変更したグラフを取得

CONSTRUCT {

?s foaf:name ?o.

} WHERE{

?s rdfs:label ?o.

} LIMIT 10

Named Graph

- トリプルの集合を区別したい
 例)データ元が違うものを区別
- トリプルでなくクアッド(4つ組)になる
- 例)データ元(dc:source)ごとにグラフがある場合に 対象リソースを主語とするトリプルを含むグラフ とそのデータ元を取得

```
SELECT DISTINCT ?g ?source WHERE {
    GRAPH ?g {
        <http://lod.ac/species/Papilio_xuthus> ?p ?o .
    }
    ?g dc:source ?source .
    ※ LODAC Species で試せる
```

SPARQL 検索テクニック





- FILTER
 - 取得したデータを条件で絞り込む
- OPTIONAL
 - 一部のグラフパターンをオプションに して完全にマッチしない場合も許容
- ORDER BY
 - 変数によって昇順もしくは降順にソート


FILTER(緯度・経度の範囲指定) 例)地図上の指定された矩形の範囲にある リソースを取得したい

s	label	lat	long
http://sgmhr.jp/place/48	共和中学校	35.566076	139.397374
http://sgmhr.jp/place/47	共和小学校	35.5610081907	139.395447555



FILTER + regex (文字列検索)

regex(文字列変数,正規表現のパターン)

例) ラベルに「小学校」という文字列を含む リソースだけを5件取得したい

SELECT * WHERE{

?s rdfs:label ?label.

FILTER regex(?label, "小学校")

}LIMIT 5

S	label	
http://sgmhr.jp/place/88	根小屋小学校	
http://sgmhr.jp/place/64	清新小学校	
http://sgmhr.jp/place/16	青葉小学校	
http://sgmhr.jp/place/21	新磯小学校	זור

OPTIONAL

グラフパターンの部分的なマッチングを許容

例)リソースのラベルは必須とし, 同一のリソースへのリンクがあればURIを取得

SELECT * WHERE{
 ?s rdfs:label ?label.
 OPTIONAL{ ?s owl:sameAs ?o. }
}
LIMIT 5 OFFSET 20

S	label	0
http://sgmhr.jp/place/12	青根小学校	undefined
http://sgmhr.jp/place/85	中野中学校	undefined
http://sgmhr.jp/place/77	鶴園小学校	undefined
http://sgmhr.jp/place/65	清新中学校	http://ja.dbpedia.org/resource/相模 原市立清新中学校

iative

ORDER BY

変数の値により昇順・降順にソート

例)リソースの緯度・経度を取得し, 緯度の昇順にソートして上位10件を取得

SELECT * WHERE{ ?s geo:lat ?lat; geo:long ?long. } ORDER BY ?lat LIMIT 10



s	lat	long	
http://sgmhr.jp/place/21	35.5036867935	139.383952601	
http://sgmhr.jp/place/103	35.5061564496	139.404596688	
http://sgmhr.jp/place/66	35.5087358195	139.409182983	
http://sgmhr.jp/place/67	35.5118057597	139.400254801	tis
http://combr.ip/place/106	95 5109720160	130 /03650811	UK

SPARQL 1.1



SPARQL 1.1での変更点(抜粋)

- 作成・更新・削除用の句
 - INSERT / UPDATE / DELETE / LOAD / DROP
- 結果形式の追加
 - JSON / CSV / TSV
- 機能追加
 - **プロパティパス / 集約 /** 条件 / 数値関数 / 文字列関数 / サブクエリ / Federated **クエリ** など
- 注意点
 - 運用中のエンドポイントは対応の度合いに差がある
 - 特にFederatedクエリやBINDが使えないところも



プロパティパス

 複数のプロパティを経由した先の オブジェクトを取得したい

例)リソースの上位カテゴリのラベルを取得

PREFIX dbpedia-ja: <http://ja.dbpedia.org/resource/> PREFIX prop-ja: <http://ja.dbpedia.org/property/> PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>

SELECT DISTINCT ?imgurl WHERE { dbpedia-ja:東京都 prop-ja:花 ?flower.



?flower dbpedia-owl:thumbnail ?imgurl.

※ <u>DBpedia Japanese</u>で試せる



プロパティパス

 複数のプロパティを経由した先の オブジェクトを取得したい

例)リソースの上位カテゴリのラベルを取得

PREFIX dbpedia-ja: <http://ja.dbpedia.org/resource/>
PREFIX prop-ja: <http://ja.dbpedia.org/property/>
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?imgurl WHERE {
 dbpedia-ja:東京都
 prop-ja:花/dbpedia-owl:thumbnail ?imgurl.
}
<u>
 X DBpedia Japanese</u>で試せる

※/(スラッシュ)で複数のプロパティをつなぐ

集約:COUNT

• COUNT / SUM / MIN / MAX / AVG / GROUP_CONCAT / SAMPLE など

```
例)小学校をラベルに含むリソースの数を取得
```

```
SELECT (COUNT(?s) AS ?count) WHERE{
?s rdfs:label ?label.
FILTER regex(?label, "小学校")
}
```

count

72

※ ?s をCOUNTで数えた数値を ASで ?count に置きかえて, SELECTで取得



集約:COUNT + GROUP BY

• GROUP BY でトリプルをグループ化してカウント

例)地域区分(区)ごとに小学校の数をカウント

PREFIX sgmhr: <http://sgmhr.jp/> SELECT ?cat (COUNT(?s) AS ?count) WHERE{ ?s rdfs:label ?label; sgmhr:areaCategory ?cat. FILTER regex(?label, "小学校") } GROUP BY ?cat

cat	count
中央区	21
緑区	27
南区	24

※ 地域区分 ?cat に基いて変数の値の組み合わせをグループ化 (?catの数だけグループができる)し、各グループが含む 組合せの数を COUNTで数えて、ASで?countに置き換え

Federatedクエリ

• 複数のSPARQLエンドポイントを使った問合せ

PREFIX skos: <http://www.w3.org/2004/02/skos/core#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX prop-ja: <http://ja.dbpedia.org/property/> SELECT * WHERE{ <http://lod.ac/id/3172> skos:prefLabel ?label. SERVICE <http://ja.dbpedia.org/sparql> { ?s rdfs:label ?label; rdfs:comment ?comment; prop-ja:愛称 ?nickname. } LIMIT 10 With the state of the state of

label	Φ	s ¢	comment	¢	nicknams
横浜美術館 @ja		<u>http://ja.dbpedia.org/resource/横浜美術館</u>	橫浜美術館	(よこはまびじゅつかん)は、横浜市みなとみらい地区(36街区)にある美術館。 @ja	ヨコピ @ja

SPARQLの練習



SPARQLを練習してみましょう

- 練習場所: DBpedia Japanese
- <u>http://ja.dbpedia.org/sparql</u>

DBpedia Japanese	ホーム	SPARQL Endpoint	Faceted Browser	SPARQL	ダウンロード
DBpediatWikipedia を対象としています	<mark>と ciaanese</mark> anese bbpedia	抽出してLOD (Linked Japanese の目的は、V	Open Data)として公開 Vikipedia日本語版をタ	削するコミュニ 対象としたDBp	- ティブロジェクトです。本家のDBpediaは主にWikipedia英語版 eediaを提供することです。
リソース係 - 東京都 - 森薫 - 国立情報学研3 - 情報・システー - 新宿駅 - 建長寺 - サザンオール3	〕 ^{兄所} ム研究機構 スターズ			SPARC SPARC 東京 SELLE: WHER	RQL 例 2L Endpoint 我都 CT DISTINCT * E { tp://ja.dbpedia.org/resource/東京都> ?p ?o .
 鶴岡八幡宮 国道行号 初音ミク ももいろクロー 知床半島 平清盛 姫路城 レオナルド・ク 	ーバーZ ダ・ヴィンヲ	<u>.</u>		} SPARC ロッ も)	DL結果 クク音楽のリスト (もしあれば画像uri CT DISTINCT ?label ?depiction

Virtuoso SPARC	2L Query Editor
Default Data Sat Nama	(Creach IRI)
berault Data Set Name	(Graph IN)
http://ja.ubpedia.org	
Query Text	
select distinct *	where { <http: ja.dbpedia.org="" resource="" 東京都=""> ?p ?o . }</http:>
Security restrictions of this serve	er do not allow you to retrieve remote RDF data, see <u>details</u>)
Results Format.	0 milliseconds (values less than 1000 are impred)
Execution timeout:	Strict checking of void variables
Options:	
The result can only be sent bac	:k to browser, not saved on the server, see <u>details</u>)
Run Query Reset	
Virtuos	Copyright © 2014 <u>OpenLink Software</u> o version 07.00.3203 on Linux (x86_64-unknown-linux-gnu), Single Server Edition



SPARQL 穴埋め問題 (1)

• ラベルに「関内」を含むリソースを取得する

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?s WHERE {
    ?s ?label.
FILTER regex(_______, "関内")
}LIMIT 10
```



SPARQL 穴埋め問題 (2)

• 横浜市の「木」を示すリソースを取得する

```
PREFIX dbpedia-ja: <http://ja.dbpedia.org/resource/>
PREFIX prop-ja: <http://ja.dbpedia.org/property/>
SELECT ?tree WHERE {
dbpedia-ja:横浜市 ?tree.
}
LIMIT 10
```

くヒント> dbpedia-ja:横浜市を主語とした全トリプルを取得し 木にあたる述語がないか調べてみる



SPARQL 穴埋め問題(3)

 日本の市の名前とその花の一覧を取得 (完璧なリストでなくてOK)

PREFIX prop-ja: <http://ja.dbpedia.org/property/> SELECT * WHERE {

くヒント> dbpedia-ja:横浜市を主語とした全トリプルを取得し 日本の市を表す述語・目的語の組がないか調べる いたにNed Open Data Initiative

SPARQL 応用問題

最も市の花に指定されている数が多い
 花のリソースとその指定数のランキングを作成

PREFIX prop-ja: http://ja.dbpedia.org/property/

SELECT

WHERE {

穴埋め問題(3)の解答を利用



- プロパティと値の組合せを値でグループ化してカウント
- カウントした数で降順に並べ替える

Linked Open Data Initiative

SPARQLを使って アプリケーションを作る



LODを使ってアプリを作る

- RDF/SPARQLのためのライブラリ
- SPARQL検索やRDFのパースなど

–Ruby: RDF.rb

- Python: SPARQLWrapper
- -PHP: sparqllib.php
- -Java: ARQ for Jena



LODAC Location http://lod.ac/apps/lodac_location



(35.57836874364153, 139.41316600297841), (35.56440595557225, 139.38741679643545)

1 渕野辺ひばり幼稚園 (約0.1km) [LinkedGe 2 渕野辺ひばり幼稚園(約0.1km)[LODAC] 3 神奈川県相模原市淵野辺四丁目(約0.2km) 4 中央シェル石油販売 ふちのべSS (約0.2kn 5 クリエイト 相模原淵野辺本町店 (約0.2km 6 セイムス 淵野辺店 (約0.2km) [LinkedGeoI 7 セイムス 淵野辺店 (約0.2km) [LinkedGeoI 8 相模原消防署淵野辺分署(約0.3km) [LOD. 9 神奈川県相模原市淵野辺本町二丁目(約0. 10 夢庵 淵野辺店 (約0.3km) [LinkedGeoData 11 相模原消防署淵野辺分署(約0.3km) [Linl 12 相模原市立幸町児童館 (約0.3km) [LODA 13 淵野辺郵便局 (約0.3km) [LODAC] 14 淵野辺郵便局 (約0.3km) [LinkedGeoData] 15 相模原市立幸町児童館 (約0.3km) [Linked 16 華屋与兵衛 淵野辺店 (約0.3km) [LinkedC 17 コクピット 相模原 (約0.3km) [LinkedGec 18 コクピット 相模原 (約0.3km) [LinkedGec 19ファミリーマート 木藤渕野辺四丁目店(20 市立淵野辺小学校(約0.3km) [LinkedGeo] 21 市立淵野辺小学校(約0.4km) [LODAC] 22 市立大野北小学校(約0.4km) [LODAC] 23 青山学院大学(国際政治経済学部)(約(24 青山学院大学(理工学部)(約0.4km)[L 25 青山学院大学(経済学部)(約0.4km)[L 26 青山学院大学 (経営学部) (約0.4km) [L 27 青山学院大学(文学部) (約0.4km) [LOI 28 青山学院大学(法学部)(約0.4km)[LOI inked Open Data Initiative. Linked Open Data.jp 🛧 🛧 🛧 🛧

LODAC LocationのSPARQL

• DBpedia, DBpedia Japanese, LinkedGeoDataは 同じクエリで問合せ

SELECT DISTINCT ?link ?title ?lat ?long WHERE { ?link rdfs:label ?title; geo:lat ?lat; geo:long ?long. FILTER (?lat > 矩形の南西の点の緯度 && ?lat < 矩形の北東の点の緯度 && ?long > 矩形の南西の点の経度 && ?long < 矩形の北東の点の経度 && lang(?title) = "ja" ※ Prefix省略

APIと違って問合せ先が違って もクエリが同じ場合も

Yokohama Art Spot http://lod.ac/apps/yas





◇ c1924(大正13/四) 日本面 編本着色 軸/138(0/51.7cm)

上村松園

Yokohama Art SpotのSPARQL (一部)

• LODAC Museumから対象施設の所蔵品情報を取得

SELECT ?link ?title ?creator ?created ?genre ?material ?size WHERE {

?lodacuri dc:references ヨコハマ・アートLODでの施設のURI;

lodac:isProviderOf ?link.

?link rdfs:label ?title;

dc:references ?workRef.

?workRef lodac:genre ジャンル名;

dc11:creator ?creator;

dc:medium ?material;

dc:extent ?size.

OPTIONAL{ ?workRef dc:created ?created; }



_INKED Open Data Initiative

SPARQLを使ったLOD可視化(1)

• Sgvizler

http://dev.data2000.no/sgvizler/

- SPARQLで取得した結果を 様々なグラフで表示
- Google Visualization APIを利用
- SPARQL Timeliner

http://uedayou.net/SPARQLTimeliner/

 – SPARQLでタイトル,位置, 時間を取得すると, 年表と地図で可視化





SPARQLを使ったLOD可視化(2)

• Gephi

http://dev.data2000.no/sgvizler/

+ SemanticWebImport

https://marketplace.gephi.org/plugin/semanticwebimport/

- グラフ可視化アプリケーション
- SemanticWebImportの プラグインを入れることで SPARQLで取得したグラフを CONSTRUCTで整形して 可視化できる



DBpediaから映画に関する グラフを可視化 、 Linked Open Data Initiative

SPARQLを もっと知りたい もっと試したい



説明で用いたSPARQLエンドポイント

- 相模原市の小・中学校など(SparqlEPCU上)
 - <u>http://lodcu.cs.chubu.ac.jp/SparqIEPCU/</u> project.jsp?projectID=sagamihara_school
- DBpedia Japanese (日本語WikipediaのLOD)

– <u>http://ja.dbpedia.org/sparql</u>

• LODAC Museum (国内の博物館情報のLOD)

<u>http://lod.ac/sparql</u>

- LODAC Species (生物多様性情報のLOD)
 - http://lod.ac/species/sparql



SPARQLを勉強する

- W3C勧告(英語ですが日本語訳も検索すればあります)
 - SPARQL 1.0 : <u>http://www.w3.org/TR/rdf-sparql-query/</u>
 - SPARQL 1.1 : <u>http://www.w3.org/TR/sparql11-query/</u>
- 本
 - Learning SPARQL(英語)
 <u>http://www.amazon.co.jp/dp/1449306594</u>
- SPARQLエンドポイント集



- datahub api/sparql
 <u>http://datahub.io/dataset?q=&res_format=api%2Fsparql</u>
- SPARQL Endpoints Status <u>http://sparqles.okfn.org/availability</u>
- サンプルクエリ集
 - LODAC SPARQLスニペット(LODAC Museumで試せます) http://lod.ac/wiki/SPARQLhttp://lod.ac/wiki/SPARQLスニペット



ぜひ一度ご自分で試してみてください

表データからLODを作成

– Open Refine + RDF Refine

作ったLODを公開

- SparqIEPCU

- 公開されたLODに問合せ
 - -SPARQL (RDFクエリ言語)



次回はいよいよ最終回!

第6回 7/7 (月)

▼LODシステム実践紹介(小林巌生)

LODを採用したシステム設計について実際の二つの例をもとに解説し ます。ヨコハマ・アート・LODでは、公益財団法人横浜市芸術文化振 興財団が管理する複数のウェブシステムをLODで統合し、データの共 有を実現しています。SPARQLエンドポイントを活用したアプリの開 発などデータの応用例も広がっています。かなざわ育なび.netでは市 役所内の複数の原課によってそれぞれ管理されていたデータをLODに よって統合して子育て世帯向けの情報提供サービスを実現しました。 また、地域でのLODの実践について今後の展望についても紹介します。

