

合同会社 緑 I T 事務所

Midori IT Office, LLC

LODとSPARQL入門(2)

この記事は1年以上前に書かれました。
内容が古くなっている可能性がありますのでご注意ください。

※当サイトを <https> に変更したため、<http> で運用されているSPARQLエンドポイントとの接続でエラーになり、チャートが表示されなくなっています。

リンクト・オープン・データ (LOD : Linked Open Data) とそのクエリ言語SPARQL (スパークル) 入門の第2回では、経済産業省がLODとして公開しているデータセットを使用します。SPARQLエンドポイントは[こちら](#)になります。
今回はさらに、WordPressでSgvizlerを使うためのプラグインを利用してクエリ結果をビジュアライズしようと思います (そのため、データを提供しているサーバがダウンしている時にはチャートが正しく表示されません)。

はじめに使うデータセットは、平成23年度エネルギー消費統計調査です。元になったExcelファイルは[こちら](#)にあります。
Excel表をLODにする方法は、[こちら](#)に解説があります。平たく言えば、Excel表で数値の入ったセルの1つ1つに対して、それは何の数値なのかというアノテーションを付けるという方法です。

そこで、[SPARQLエンドポイント](#)で以下のクエリを発行すると、平成23年度エネルギー消費統計調査のデータセット<<http://datameti.go.jp/lod/energy-shouhi-toukei/h23-2-2-B>>のセルに対してどのようなアノテーションが付けられているかが分かります。

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX toukei: <http://datameti.go.jp/lod/energy-shouhi-toukei/>

SELECT DISTINCT ?p WHERE {
  ?s qb:dataSet toukei:h23-2-2-B;
  ?p ?o.
}
```

実行すると以下の結果が得られます。

```
P
http://www.w3.org/1999/02/22-rdf-syntax-ns#type
http://datameti.go.jp/scheme/energy-census-qbproperty/amountOfEnergy
http://datameti.go.jp/scheme/energy-census-qbproperty/refEnergy
http://datameti.go.jp/scheme/energy-census-qbproperty/refSangyoBunrui
http://purl.org/linked-data/cube#dataSet
http://purl.org/linked-data/cube#measureType
```

エネルギーの種類 (refEnergy) や産業分類 (refSangyoBunrui) 毎にエネルギー消費量 (amountOfEnergy) が入っています。

それでは、エネルギーの種類毎の消費量を取得してみましょう。

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX toukei: <http://datameti.go.jp/lod/energy-shouhi-toukei/>
PREFIX qbproperty: <http://datameti.go.jp/scheme/energy-census-qbproperty/>
PREFIX energy: <http://datameti.go.jp/scheme/energy/2013/>

SELECT ?slabel SUM(?e) AS ?energy WHERE {
  ?cell qb:dataSet toukei:h23-2-2-B ;
  qbproperty:refEnergy ?type ;
  qbproperty:amountOfEnergy ?e .
  ?type rdfs:label ?slabel .
}
GROUP BY ?slabel
ORDER BY DESC(?energy)
```

Sgvizlerで棒グラフにすると以下のようになります。

[WP Sgv] Failed to draw chart. Maybe server error.

「燃料」がダントツで「非石油系燃料」と「石油系燃料」がそれに続いています...なんか変ですね。実は、元のExcelファイルを見ればすぐに分かるのですが、エネルギーの種類は階層化されており、「非石油系燃料」や「石油系燃料」は「燃料」の内訳なのです。さらに、「石油系燃料」は「ガソリン」や「原油」をなど含んでいます。分類階層毎に見ないと意味がありません。

そこで、最上位の分類で集計することになります。

qbproperty:refEnergyの中で、skos:topConceptOf energy:EnergyConceptSchemeという述語と目的語を持っているのが、最上位のエネルギー分類です。

そこで、以下のようなSPARQLクエリを発行し、

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX toukei: <http://datameti.go.jp/lod/energy-shouhi-toukei/>
PREFIX qbproperty: <http://datameti.go.jp/scheme/energy-census-qbproperty/>
PREFIX energy: <http://datameti.go.jp/scheme/energy/2013/>

SELECT ?slabel SUM(?e) AS ?energy WHERE {
  ?cell qb:dataSet toukei:h23-2-2-B ;
  qbproperty:refEnergy ?type ;
  qbproperty:amountOfEnergy ?e .
  ?type skos:topConceptOf energy:EnergyConceptScheme .
  ?type rdfs:label ?slabel .
}
GROUP BY ?slabel
ORDER BY DESC(?energy)
```

今度は円グラフにしてみます。

[WP Sgv] Failed to draw chart. Maybe server error.

3つの大分類毎のエネルギー消費量を可視化することができました。

次は、産業分類毎に見てみます。

以下のようなSPARQLクエリを発行します。

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX toukei: <http://datameti.go.jp/lod/energy-shouhi-toukei/>
PREFIX qbproperty: <http://datameti.go.jp/scheme/energy-census-qbproperty/>
PREFIX energy: <http://datameti.go.jp/scheme/energy/2013/>

SELECT DISTINCT ?slabel SUM(?e) AS ?energy WHERE {
  ?cell qb:dataSet toukei:h23-2-2-B ;
  qbproperty:refEnergy energy:燃料 ;
  qbproperty:amountOfEnergy ?e ;
  qbproperty:refSangyoBunrui ?sangyoBunrui .
  ?sangyoBunrui skos:broader ?sangyo .
  ?sangyo rdfs:label ?slabel .
}
GROUP BY ?slabel
ORDER BY DESC(?energy)
```

qbproperty:refEnergy energy:燃料；

で対象を「燃料」に限定し、

qbproperty:refSangyoBunrui ?sangyoBunrui .

と

?sangyoBunrui skos:broader ?sangyo .

で産業の大分類毎の値を対象とします。そして

GROUP BY ?slabel

と

SUM(?e) AS ?energy

で産業の大分類毎に集計します。

今度は3D円グラフにしてみます。

[WP Sgv] Failed to draw chart. Maybe server error.

製造業が圧倒的に大量の燃料を消費しています。運送業の燃料消費は第10位で、異常に少ないように思います。そこで元のExcelファイルを見たところ、「営業用の車、バス、鉄道等のエネルギーは含まれません」との注釈がありました。

同様に、電力の消費量を産業分類毎に見てみます。

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX toukei: <http://datameti.go.jp/lod/energy-shouhi-toukei/>
PREFIX qbproperty: <http://datameti.go.jp/scheme/energy-census-qbproperty/>
PREFIX energy: <http://datameti.go.jp/scheme/energy/2013/>

SELECT DISTINCT ?slabel SUM(?e) AS ?energy WHERE {
  ?cell qb:dataSet toukei:h23-2-2-B ;
  qbproperty:refEnergy energy:電力 ;
  qbproperty:amountOfEnergy ?e ;
  qbproperty:refSangyoBunrui ?sangyoBunrui .
  ?sangyoBunrui skos:broader ?sangyo .
  ?sangyo rdfs:label ?slabel .
}
```

```
}  
GROUP BY ?slabel  
ORDER BY DESC(?energy)
```

結果は以下のようになります。

[WP Sgv] Failed to draw chart. Maybe server error.

燃料と比べると製造業の比率が低下し、卸売業・小売業が第2位になりました。

LODとSPARQL入門 [1](#) [2](#) [3](#) [4](#)

カテゴリー: オープンデータ | タグ: LOD, SPARQL | 投稿日: 2014年3月15日 [<https://midoriit.com/2014/03/lod%e3%81%a8sparql%e5%85%a5%e9%96%802.html>] | 投稿者: 小池隆
