

SPARQL APIのご紹介

2014年3月20日

先端IT活用推進コンソーシアム
クラウド・テクノロジー活用部会
荒本 道隆

SPARQLとは

SPARQL ("スパークル"と発音[1])はRDFクエリ言語の一種である。その名称は再帰的頭字語になっており、**SPARQL Protocol and RDF Query Language**の略。RDFクエリ言語とは、[Resource Description Framework](#)で記述されたデータを検索/操作するコンピュータ言語である。SPARQLは[World Wide Web Consortium](#) (W3C)の[RDF Data Access Working Group](#) (DAWG)による標準化作業が行われている。W3C勧告に至る過程で、2006年4月に勧告候補としてリリースされたが、2つの問題により2006年10月に草案に戻された[2]。2007年6月、SPARQLは再び勧告候補となった[3]。2008年1月15日、SPARQLはW3C勧告となった[4]。SPARQLはクエリのパターンとして、[論理積](#)、[論理和](#)、その他のパターンを指定可能である[5]。

ウィキペディアより

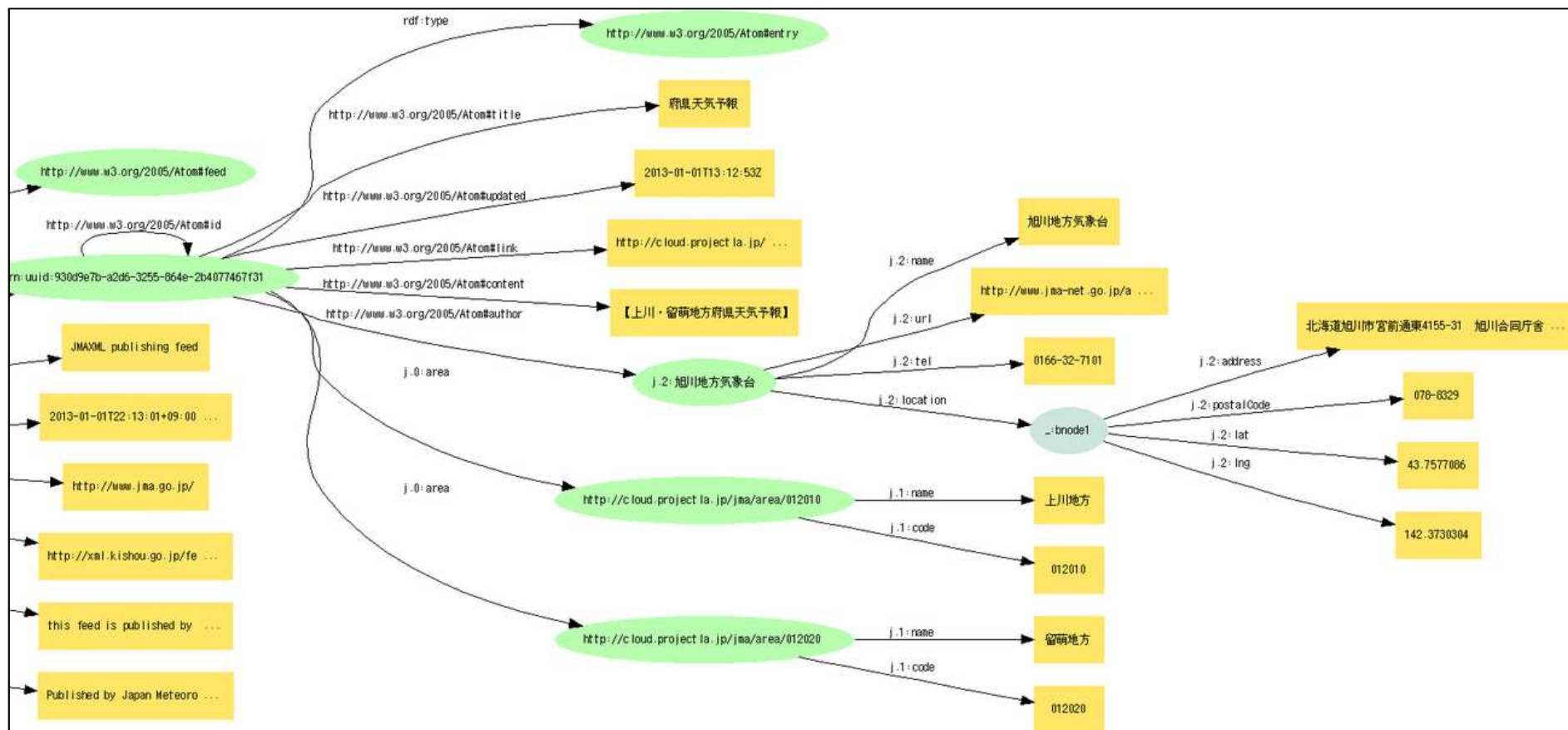
- まとめると
 - RDBに対するSQLのようなもの
 - RDFというデータセットに対して検索、集計、フィルタリングなどが行える

気象庁防災情報XMLとRDFの関係



- 気象庁XML ← これをRDF化するのは、とても大変
 - 気象庁は、これまで、**気象警報、津波警報、地震情報等、それぞれの防災情報**毎に情報の性質・利用形態などを考慮し、それぞれの情報で個別の、気象庁独自の電文形式(フォーマット)を作成してきました。この方式は、防災情報の種類が少なく、情報の伝達がFAXや低速の通信回線の時代はそれぞれの情報に適していましたが、高度にICT化された現在社会において、より詳細で高度化された防災情報をより効果的に活用していただくために、新たな防災情報の提供様式を検討すべきと考え、「気象庁防災情報**XMLフォーマット**」を策定することとし、平成23年5月12日より運用を開始しました。
<http://xml.kishou.go.jp/>
- feed ← これをRDF化 + エリア情報 + 気象台の位置情報
 - 気象庁 → Google PubSubHubbub経由でfeedを配信
 - **概要と気象庁防災情報XMLのURL**が記述されている
 - そのURLにアクセスして、気象庁XML本体を取得
 - 1つのfeedの中に、複数のentryが入っている

RDFの構造



<http://www.kanzaki.com/works/2005/rap/graph-check> でサンプルRDFを可視化

配信数の詳細 (2013/01/01 ~ 2014/03/11)



"府県天気概況"	94760	"震源に関する情報"	266
"府県天気予報"	85292	"地方高温注意情報"	259
"気象警報・注意報"	54302	"火山の状況に関する解説情報"	240
"府県週間天気予報"	48426	"府県潮位情報"	230
"気象特別警報・警報・注意報"	24312	"全般気象情報"	217
"地方海上警報"	14097	"季節観測"	211
"地方海上予報"	10454	"スモッグ気象情報"	193
"地方週間天気予報"	9619	"府県天候情報"	186
"府県気象情報"	8431	"地方3か月予報"	143
"紫外線観測データ"	7359	"全般台風情報"	122
"震源・震度に関する情報"	2686	"全般台風情報(詳細)"	98
"特殊気象報"	2407	"記録的短時間大雨情報"	77
"噴火に関する火山観測報"	2204	"地方天候情報"	71
"地方気象情報"	2163	"地方潮位情報"	68
"台風解析・予報情報(3日予報)"	1811	"全般1か月予報"	63
"全般海上警報(定時)"	1746	"地方暖・寒候期予報"	33
"生物季節観測"	1668	"全般潮位情報"	30
"竜巻注意情報"	1548	"津波情報"	15
"府県高温注意情報"	1412	"全般天候情報"	14
"土砂災害警戒情報"	1002	"東海地震観測情報"	14
"指定河川洪水予報"	833	"全般3か月予報"	13
"地方1か月予報"	693	"地震の活動状況等に関する情報"	10
"全般台風情報(定型)"	599	"津波警報・注意報・予報"	9
"全般週間天気予報"	512	"噴火警報・予報"	8
"府県海氷予報"	403	"気象特別警報報知"	7
"異常天候早期警戒情報"	401	"津波警報・注意報・予報a"	7
"震度速報"	372	"津波情報a"	6
"台風解析・予報情報(5日予報)"	320	"全般スモッグ気象情報"	5
"全般海上警報(臨時)"	301	"顕著な地震の震源要素更新のお知らせ"	5
		"JMAXML publishing feed"@ja	4
		"全般暖・寒候期予報"	3
		"火山現象に関する海上警報・海上予報"	3

SPARQLの画面から参照可能

実際の配信数の概要



- 2013/01/01 ~ 2013/12/31 (1年間)
 - feed
 - 572,595回 (平均1,568回 / 日)、重複含む
 - 2,444MByte / 年 (平均6.69MByte / 日)
 - 気象庁防災情報XML
 - 323,212件 (平均885件 / 日)
 - 6,399MByte / 年 (平均17.53MByte / 日)

SPARQLでできること - 1

- 欲しい形で結果を取得
 - 結果のカラム名や形式を自分で指定できる
 - 利用できる集計関数
 - COUNT, SUM, MIN, MAX, AVG, GROUP_CONCAT, SAMPLE
 - サイトに掲載しているサンプル
 - Tripleの取得(元データのダンプ)
 - エリア名とコードのリスト
 - タイトルで集計 ← デフォルトでこの行が有効になっています
 - 特定のタイトルのものの中身を参照
 - 「雪」に関するコンテンツがどの气象台からいくつ出ているか？
 - "全般週間天気予報"のうち、2014/01/01に配信されたもの
- アプリへの組み込みが容易
 - HTTPでGETするだけ
 - エンドポイント: [http://api.aitc.jp/ds/sparql?query=\(SPARQL\)](http://api.aitc.jp/ds/sparql?query=(SPARQL))
 - 更新系は別URLなので、勝手に更新される心配がない

SPARQLでできること - 1

- 2014/02/14 (関東で大雪だった日) に、
 - 「雪」に関するコンテンツが、
 - どのエリアに何回出たか？

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
PREFIX atom: <http://www.w3.org/2005/Atom#>
```

```
PREFIX jma: <http://cloud.projectla.jp/jma/>
```

```
PREFIX area: <http://cloud.projectla.jp/jma/area#>
```

```
SELECT ?area ?areaname (COUNT(?id) as ?c) WHERE {
```

```
  ?id atom:author ?author .
```

```
  ?id atom:content ?content .
```

```
  ?id atom:updated ?updated .
```

```
  ?id jma:area ?area .
```

```
  ?area area:name ?areaname .
```

```
  FILTER(
```

```
    xsd:dateTime(?updated) >= "2014-02-14T00:00:00+09:00"^^xsd:dateTime &&
```

```
    xsd:dateTime(?updated) < "2014-02-15T00:00:00+09:00"^^xsd:dateTime &&
```

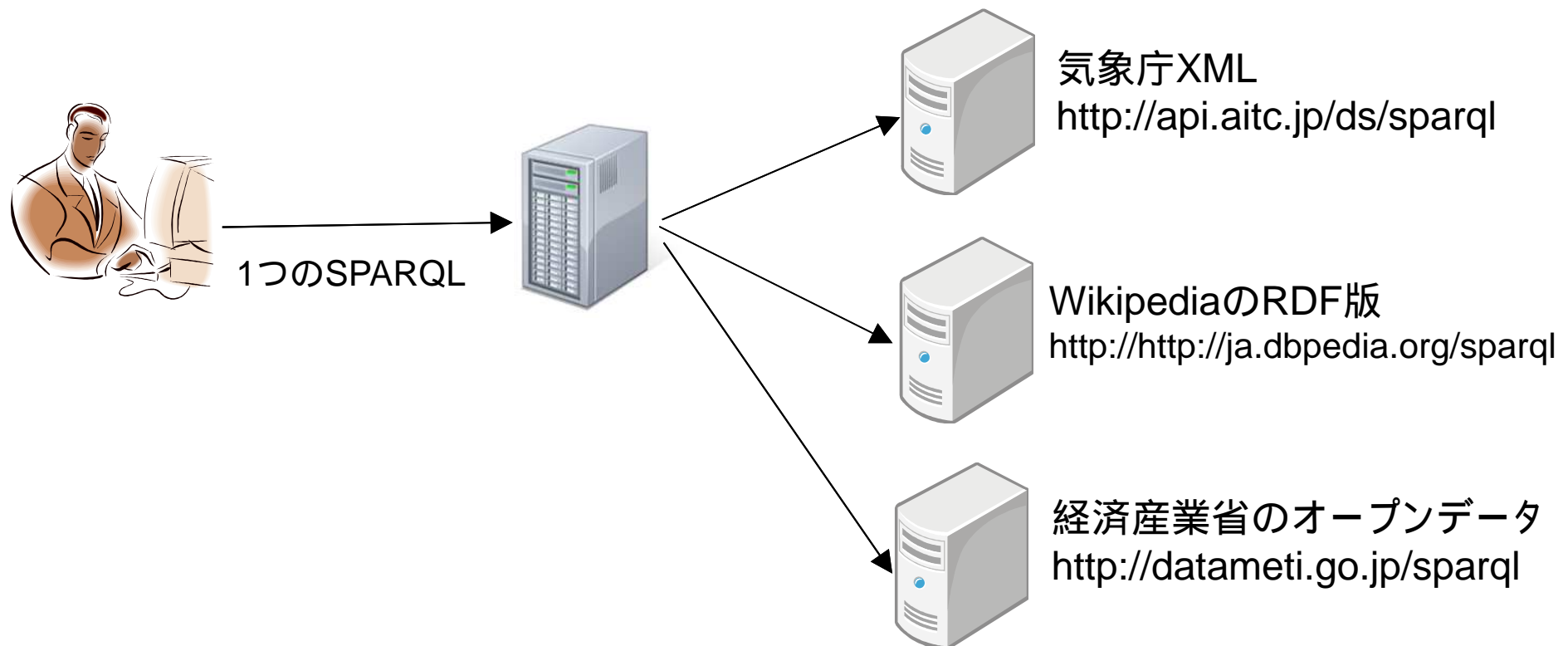
```
    REGEX(?content, "雪")
```

```
  )
```

```
} GROUP BY ?area ?areaname ORDER BY DESC(?c)
```

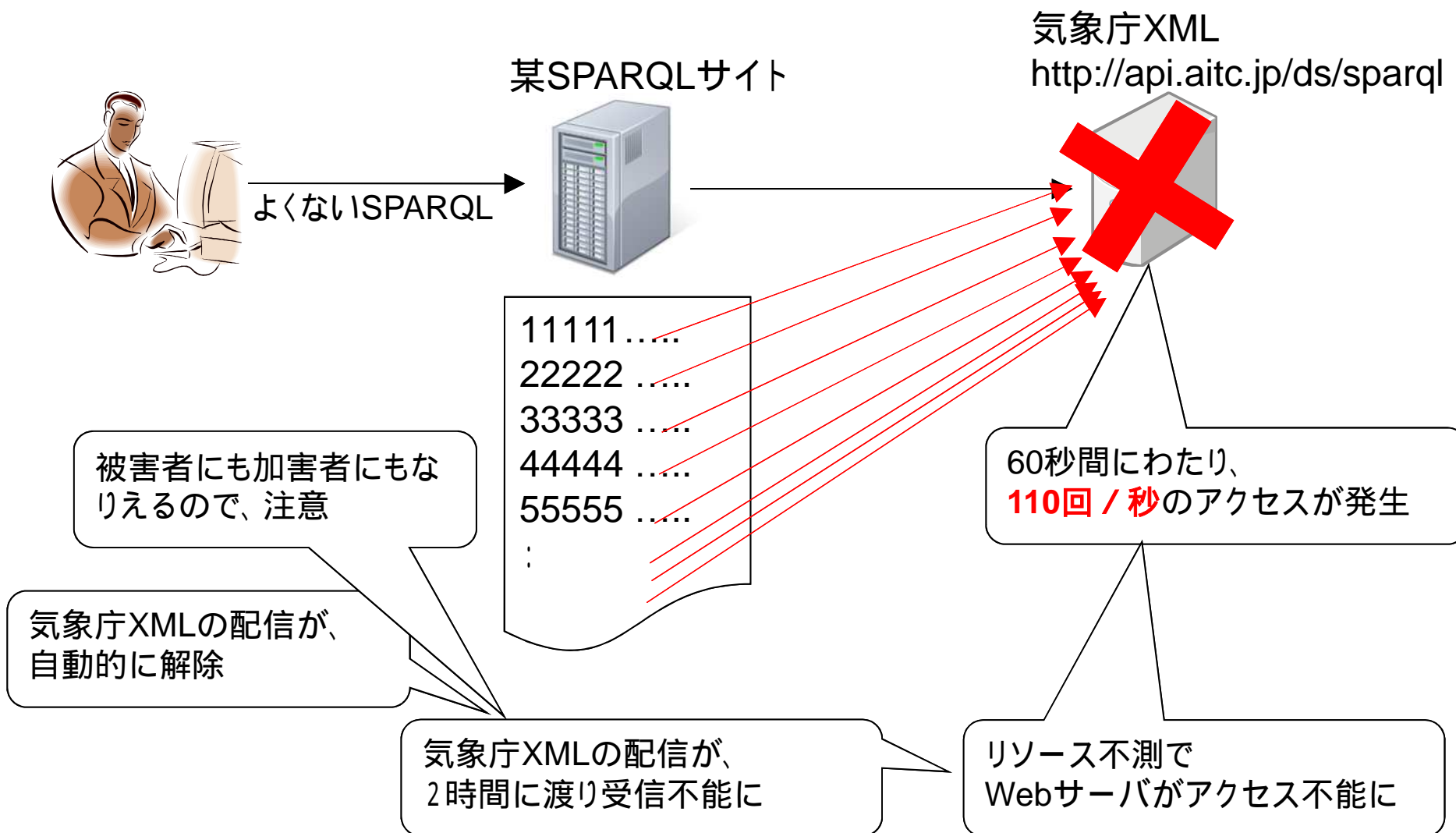

SPARQLでできること - 2

- オープンデータを使ったマッシュアップ
 - SPARQL 1.1 Federated Query
 - 1つのSPARQLで、複数サイトのRDFデータを結合



SPARQLでできること - 2

- やって見たら、こうなりました



- LODでは、SPARQLが共通言語になる
 - LOD: Linked Open Data
 - SPARQLが、今のSQLのようになるのは遠くない
 - フェデレーション機能で、より便利に
 - 可視化ツールとの組み合わせで、開発が不要に？
- 気象庁XMLをRDF化する時代が来るかも
 - XBRLをRDF化する研究なども行われている
 - XBRL: eXtensible Business Reporting Language
 - 企業の決算報告を1つの巨大なXMLにしたもの
 - 他のLODとの親和性が、より高まる
- 今からSPARQLに慣れ親しんでおきましょう

- サイトは、AITCの会費で運用しています
 - そのため、商用利用は遠慮してもらっています
 - 商用利用したい場合は、気象業務支援センターへ
 - CD-ROM, DVDによる販売
 - オンラインによるデータ配信
 - 2台のサーバで頑張ってます
 - RESTとSPARQLのデータは、受信モレがあります
 - 時々、高負荷になり、応答に時間がかかります
 - 時々、止まるかも知れません
- 何かあれば、 staff@aitc.jp まで