グーグルマップを用いたRDF/オントロジーのSPARQL検索サイトの構築
プロジェクト52d報告

年岡晃一　　鈴木裕利

中部大学 工学部 情報工学科

１．はじめに

インターネットは巨大なデータベースだとして、世の中の既存の書籍、画像、動画を含むあらゆるデータ、情報そしてこれから作られるデータも併せて格納するべく様々な努力が進行中である。格納された情報は分析され、集計され、新しい情報となってまた格納される。増大していくデータ及び情報はより整理された形式で取り出すことが出来なければならない。また多くの情報処理専門外の人達も参画できるようこれらの情報処理がより容易にデータの格納及びデータの取り出しが出来なければならない。これらを実現すべくセマンティック Web の試みがされて来た。それは言ってみれば情報の作り手が整理された形式で且つグローバルな合意を得た語彙をベースに情報を格納していくことであった。しかしグローバルな合意を形成していくことはしばしば難しいこともあり普及には至っていない。しかしながらここに至ってセマンティック Web の発展形として LOD という取組みが始まっている。

２．LOD としての新たな展開

LOD(Linked Open Data)とは

セマンティック Web は、W3C 主導の下にインターネット上にデータを機械が取り扱い可能に成るよう厳密な意味処理を目指していたものだが、LOD ではより実践的にインターネット上の個々の情報を RDF インスタンスとして公開且つ共有して行こうという提案である。そこではクラス、プロパティの制約に基づく意味処理で処理や管理を複雑化するよりも、各 Web サイト毎に得意分野データを持たせて Web 全体をより便利な形にしようとするものである。

情報の可視化

多くのデータを扱う上でそれらの情報を効率よくユーザに提示する方法として如何に大量の情報群を可視化していくかの課題がある。

従来のセマンティック Web では論理学を基盤としたものなので普遍的に成立する知識の扱いに限られることが多い。

しかし我々が日常的に扱う情報には時間情報と地理情報を含むものが多い。

インターネット上のデータの扱いでも、その情報が何時発生したか何時まで有効かどこで作られたかは重要である。

本プロジェクトでは以上のことから、図１のように各メタデータの記述に語彙・概念に加えて地理情報と時間情報を扱うことにする。これは RDF のプロパティに時間情報及び GPS の地理情報を加えるだけで比較的容易に実現可能である。

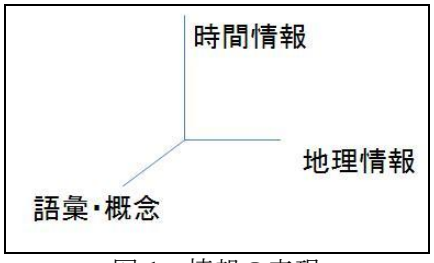

図 1．情報の表現

地理情報は携帯付属の **GPS** と地図情報提示の Google Map がインターネットで容易にアプリに導入することが出来るようになっている。

SPARQL

SPARQL[1]の検索式は関係データベースの SQL に似せたもので、端的に言うとデータモデルとして三つ組みのグラフ探索となる。実装としては内部的に関係データベースを使用し自己結合の連鎖からなるものである。

```
PREFIX base: <http://www.lod/>
SELECT ?x ?glat ?glong ?msgstr
 WHERE {
  ?x base:pos_lat ?glat.
  ?x base:pos_long ?glong.
  ?x base:msg ?msgstr
 }
```

図 2．SPARQL での検索式

その構成上関係データベースの正規化と索引を使うことに制限が生じまたテキストの前方一致、後方一致の機能は無い。

出来るだけ簡便な方法且つメタデータと内容文書の一致が自然に行われる形式が望まれる。以上をまとめると表１になる。

表１．LOD サイトの要素機能と実現方法

要素機能	実現方法
流通方式	RDF/XML/N3/CSV
格納方式	N-triples/RDB
検索方式	SPARQL
処理方式	オントロジー推論、RULE など
データ入力形式	CMS のメタデータ記述規約
利用者へのインタフェイス	キーワード検索+クラス/インスタンス統計など

３．LOD の普及に向けて設計と実装

Framework としての構成

LOD、セマンティック Web は元より Web システムに詳しくないがデータを発信或いは利用したい様々な人達にもより容易にアクセスや使用を可能にすべく種々のインタフェイスや通信、アプリケーションが提供されなければならない。図 3 に各構成間に於けるデータの流れを示す。

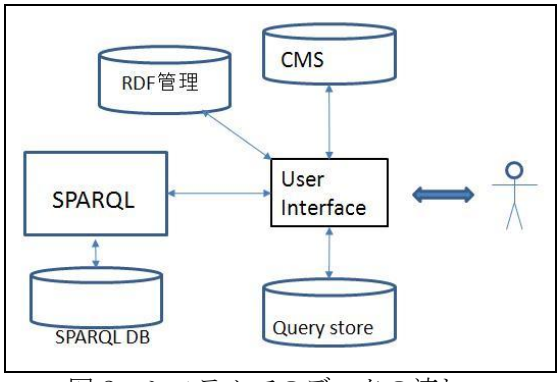

図 3．システムでのデータの流れ

今回は RDF 情報の蓄積と検索に Jena の SPARQL を使用した。RDF 情報のいわゆる永続化である

RDF store

RDF store(RDF 管理)は RDF の基本要素である三つ組み情報が主語、述語、目的語として計算機上のファイルに書き込まれたものである。
外部サイトからも容易に登録可能なように以下のデータ形式（テキスト文字列）のいずれかを送ることで RDF を登録できるようにしている。

- 表形式の記述から RDF へ
- N3 形式から RDF へ
- JSON 形式から RDF へ

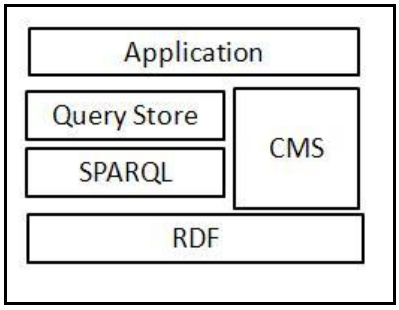

図 4．レイヤー構成

Query Store

図 5 に基本的な SPARQL を使用するユーザインタフェイスを示す。検索結果は表形式で返って来る。

図 5．構築中の SPARQL 検索サイト

通常のユーザを複雑な検索式の使用から解放することも可能にするため本プロジェクトでは Query Store と称してよく使用する検索式を関係データベースに格納させ単純な探索名呼び出しで適時利用出来るようにしている。RDF に GPS 情報が含まれていれば図 6 のように Google Map にそれが表示される。

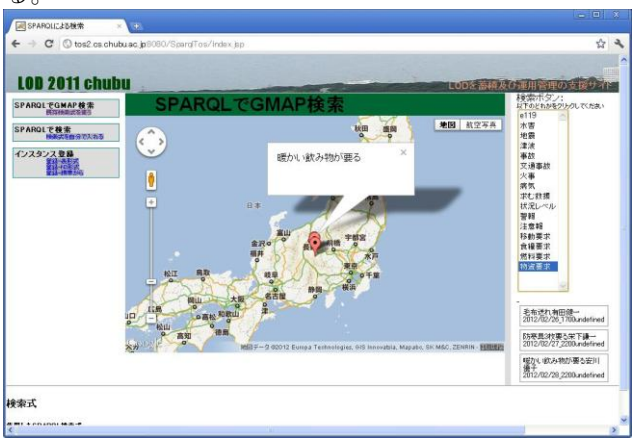

図 6．一般ユーザ向けのユーザインタフェイス

４．まとめ

本プロジェクトで進めている SPARQL と地理情報システムを用いた LOD の蓄積及び検索サイトいわゆる SPARQL End Point の設計及び実装について報告した。

政府機関も政府公開データの基盤として LOD 技術の発展に期待を寄せている。LOD 技術は新たな社会基盤として、異分野融合、学際間の連携など社会のあらゆる面における発展を促すと期待される。現在の省庁の公開データは PDF や CSV ファイルのダウンロードが可能という位のレベルで止まり、機械による自動処理の対象になるには相当の距離がある。現在はメタデータとしての LOD の登録と検索が可能であるが、従来の CMS との連携にはまだ多くの実装作業が必要である。CMS 文書データの更新と RDF として登録されたメタデータの管理作業は、使用する特殊用語の省庁の統一化や政府機関の LOD に関する政治力も必要である。今後は、インターネット上の文書更新とそのメタデータ更新との同期が可能になるべく、外部サイトの CMS でも使い易いフレームワークとしての WebAPI にまで進めて行きたい。

参考文献

[1]“SPARQL Query Language for RDF” http://www.w3.org/TR/rdf-sparql-query/

[2] “Jena” http://jena.sourceforge.net/

[3] 年岡晃一、鈴木裕利、”文書管理システム ONTDOC でのオントロジー・メタデータの利用”， 第 7 回情報科学技術フォーラム（FIT2008）L-026 Vol4 pp145-146, 2008 年 9 月.

[4] 年岡晃一、鈴木裕利、”CMS におけるセマンティック Web の利用法”、情報科学リサーチジャーナル、Vol.18、2011．3