

LOD の今後の課題と開発環境 プロジェクト 52d 報告

年岡晃一

中部大学 工学部 情報工学科

1. はじめに

次世代の情報検索システムとして提案されたセマンティック Web、言い換えると RDF による知識ネットワークは、Linked Open Data[1][2]またはオープンガバメントという形で社会的にも大きな流れになろうとしている。そこではインターネット上に在る情報・知識・概念を RDF という単純なデータ構造にすることで機械による情報の自動処理を可能にしようというものであった。しかしビジネス創出や社会システムの基盤としての役割を担う程には至っていない。ここでは LOD の今後として我々の今までの成果をまとめる中で LOD の今後解くべき課題を見て行く。

2. SparqlEPCU によるアプリ開発環境

我々が開発し公開している SparqlEPCU[3]の意味するものは、Google マップにおいて地理情報などのクラウドデータをマッシュアップでアプリケーションが容易に利用出来ていると同じことを RDF データで可能にするものである。そこでは LOD データの格納や検索を行うサーバー機能とそれを利用するアプリケーションロジックが切り分けられている。



図 1. LOD アプリ開発環境 SparqlEPCU

実際ここ一年間 SparqlEPCU に登録されたプロジェクト 90 件の半分以上の 50 件が大学外からの利用者である。図 2 に CLOUD との関係を示す。

3. LOD の今後の方向

データセットとアプリケーションの構築が進行中或いは有望視されている分野を以下に挙げる。

- オープンガバメント、国家統計、予算、地域行政
- 電子申請、法令、観光情報、学術情報
- 特許開示、リコール製品情報、詐欺・犯罪情報
- 医療、福祉、防災

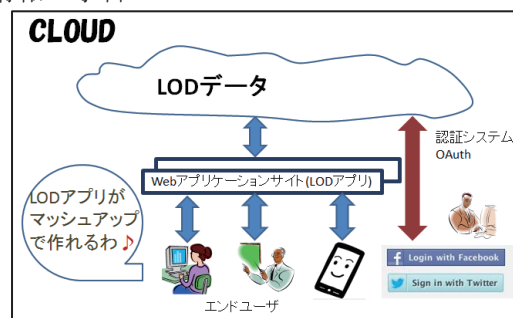


図 2. LOD と CLOUD

LOD の収集や利用形態は技術的には、解決されている。残っている問題として、情報に紐付けされるクラス（概念）や属性を社会的に認知させていく、言い換えると語彙の統合を促進する仕組みが必要と思われる。

表 1. LOD 処理方式

要素機能	実現方法
データ入力形式	CSV など（登録時）、SPARQL(検索時)
流通方式（シリアルイズ）	RDF/XML/N3/CSV/turtle
格納方式	N-triples/RDB
検索方式	SPARQL
意味処理方式	オントロジー推論、RULE など
利用者へのインタフェース	キーワード検索+クラス/インスタンス統計など
意味記述	RDF による意味ネットワーク
リソースの識別	IRI

ビッグデータとの対比

LOD はビッグデータとよく対比されるが、どちらにも求められるものとして以下がある。

1. ビジネスモデル
2. 分析手法、可視化
3. 上記を支える WEB システム

LOD はオープンなデータを基にするので収益性を求めるビジネスモデルに結び付け難い。一方、データ分析はビッグデータだけでなく LOD でも使う場合があると思われる。よく知られている統計的分析ツール R による SPARQL 検索[4]の呼出し例を図 3 に示す。

```
>install.packages("SPARQL")
>library(SPARQL)
>query<-"select * where {?s ?p ?o }limit 30"
>endpoint<-"http://lodcu.cs.chubu.ac.jp/SparqlEPCU/api/meisuitest"
>res<-SPARQL(endpoint,query)$results
>res
```

図 3. R を使った SPARQL の検索

図 4、図 5 に LOD アプリケーションの例を示す。

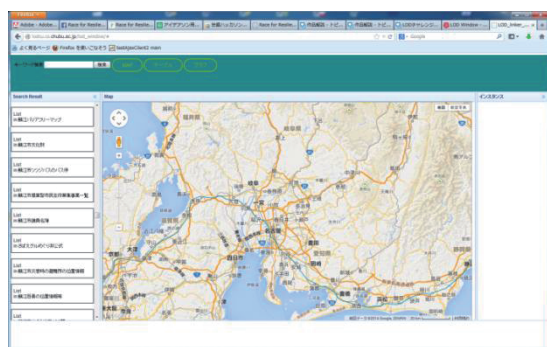


図 4. LOD window[5]



図 5. 議論展開システム[6]

4. 今後要求される要素技術

データハブ、アプリケーションハブ

Wikipedia、DBpedia、CKAN

Wikipedia は旧来の百科事典に代わった巨大な用語辞典であり且つ信頼できる情報源としてインターネット上で広く使われている。各項目はカテゴリ分けされており、そのカテゴリ階層を RDF 表現したものが DBpedia[7]である。世の中に見えるものを記述したものが Wikipedia とすれば、DBpedia はそれを構造化してコンピュータが認知できる形にしていると言える。DBpedia は SPARQL End Point として検索式を通じて情報を取りだせるようになってきている。Wikipedia と DBpedia はリソースとデータセット (RDF) の対を成す。DBpedia 以外にも数多くのデータセットが有りそれらをまとめようとするものが CKAN[8]と呼ばれるデータセット・カタログである。これらは分散している RDF データセットを統合・整理するものとしてこれから重要な存在となって行くと思われる。これらの中では RDF のクラス謂わば「概念」としてオブジェクト指向のクラスに相当するものが知識・情報の記述に役立つ場合が多い。これらはデータセットにまたがって横断的に使用されるものでそれらを探すものとして我々は LODlinker なるものを構築している (図 6 参照)。

SPARQL End Point の負荷スケールアウト

情報の源を集め検索処理を行う場所が SPARQL

End Point であるが、データが膨大になったり検



図 6. LOD linker[9]

索要求が集中した時にも対応できるように PaaS などを使って負荷を Scale Out することが必要と思われる。

データの質

収集及び検索の基盤は出来ているが、今後は格納された RDF データの鮮度・質・信憑性、情報源泉を RDF 自信で扱う簡便な方法の確立が望まれる。

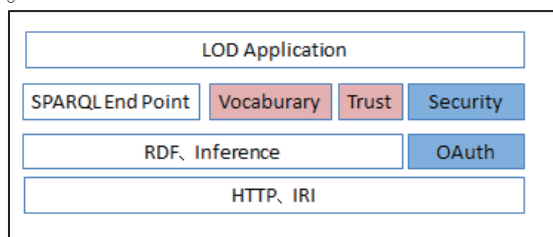


図 7. LOD レーヤーケーキ

5. まとめ

LOD 普及の鍵として、ビジネスモデルの確立、語彙の統合手続き、社会ツールとして認知確立されるなどを今後の課題として挙げた。SparqlEPCU 及び LODlinker がそれを解決するシステムとなるべく研究を進めて行く予定である。

参考文献

- [1] 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
- [2] 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 Vol14 pp145-146、2008 年 9 月。
- [3] 藤本椋也、安江翔平、年岡晃一、" LOD クラウドとしての RDF サーバー"、人工知能学会研究会資料 SIG-SW0-A1202-02、2012、10
- [4]"SPARQL Query Language for RDF" <http://www.w3.org/TR/rdf-sparql-query/>
- [5] LODwindow http://lodcu.cs.chubu.ac.jp/lod_window/
- [6] 仙北谷大雅、藤本椋也、年岡晃一、" LOD を用いたコミュニティ内の議論展開システム" 電気関係学会東海支部連合大会 2013、P01-27
- [7] DBpedia. <http://dbpedia.org/> (2012/10/01)
- [8] CKAN <http://data.linkedopendata.jp/>
- [9] LODlinker <http://lodlinker.org/>