

W3C RDF Data Access WGの 現状と今後

佐藤 宏之

日本電信電話株式会社 NTT情報流通プラットフォーム研究所
W3C RDF Data Access WGメンバ
INTAP セマンティックWeb委員会委員

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

RDF Data Access Working Group (DAWG) とは

- セマンティックWebがフェーズ2へ
 - 2004年2月にWeb Ontology Language (OWL)がW3C勧告に
 - 以降, W3Cでは実践的なアプリケーションに関する検討 (SWBPD) やルール記述言語などに活動の中心を移してきている
- 2004年2月頃, W3Cの新たな活動の1つとして誕生
 - 主にRDFのクエリ言語仕様、HTTP/SOAPによるRDFデータ取得のためのアクセスプロトコルの検討を行う
- Webとメール, 電話会議(週1回程度), Face-to-faceミーティング(2004年は年3回, 2005年も少なくとも3回を予定)による活動
 - 仕様実装の実現性と, クエリ言語の表現力や機能性とのトレードオフなどについても議論

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

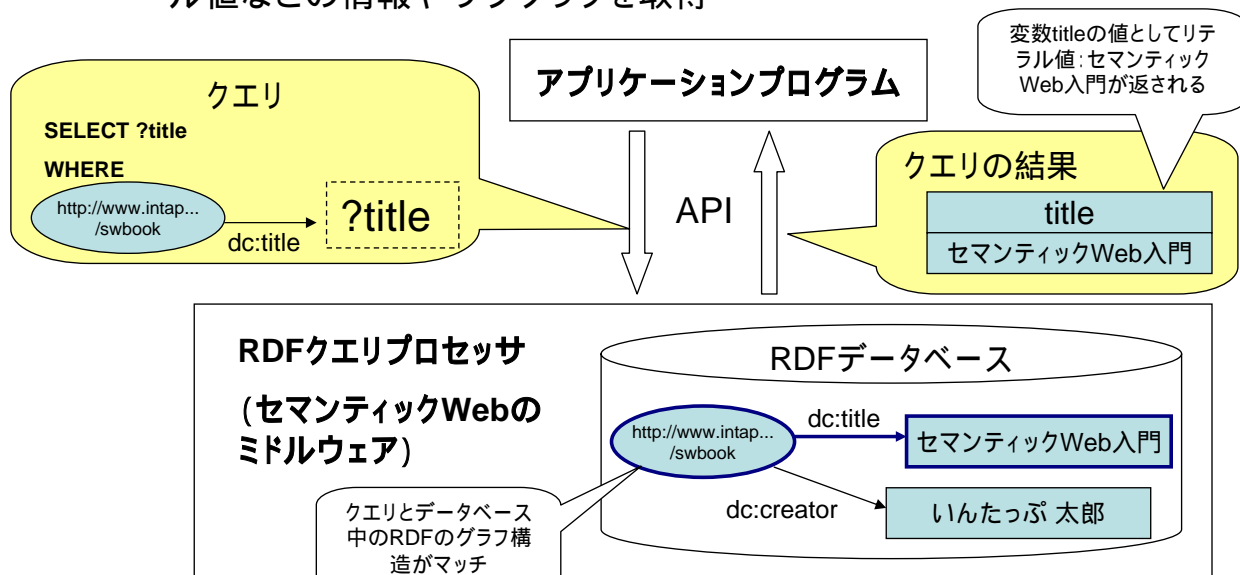
DAWGメンバ

- Chair : Dan Connolly (W3C)
- Team contact : Eric Prud'hommeaux (W3C)
- 企業 (日本からは松下電器とNTT)
 - Agfa-Gevaert N. V.
 - Jos De Roo, Dirk Colaert
 - Asemantics S.R.L.
 - Alberto Reggiori, Dirk-Willem van Gulik
 - Hewlett Packard Company
 - Andy Seaborne, Kevin Wilkinson
 - Hicks & Associates, Inc.
 - Bryan Thompson
 - Matsushita Electric Industrial Co., Ltd. (MEI)
 - Yoshio Fukushige
 - Network Inference
 - Jeff Pollock, Rob Shearer
 - Nippon Telegraph & Telephone Corp. (NTT)
 - Hiroyuki Sato
 - Profium Ltd.
 - Janne Saarela
 - Sun Microsystems, Inc.
 - Farrukh Najmi
 - Tucana Technologies, Inc.
 - Simon Raboczi, Tom Adams
- 大学および研究機関
 - Bristol, University of
 - Dave Beckett
 - Free University of Bozen-Bolzano
 - Enrico Franconi
 - Institut National de Recherche en Informatique et en Automatique (INRIA)
 - Jean-François Baget
 - Maryland Information and Network Dynamics Lab at the University of Maryland
 - Kendall Clark
 - James Hendler
 - Southampton, University of
 - Stephen Harris
- Invited Experts (招聘専門家)
 - Pat Hayes
 - Howard Katz

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

RDFのクエリ言語仕様とは

- セマンティックWebのアプリケーションで利用されるRDFのクエリプロセッサへのアクセス(入出力)を規定
 - データベースで管理されるグラフ構造のRDFデータからURIやリテラル値などの情報やサブグラフを取得



(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

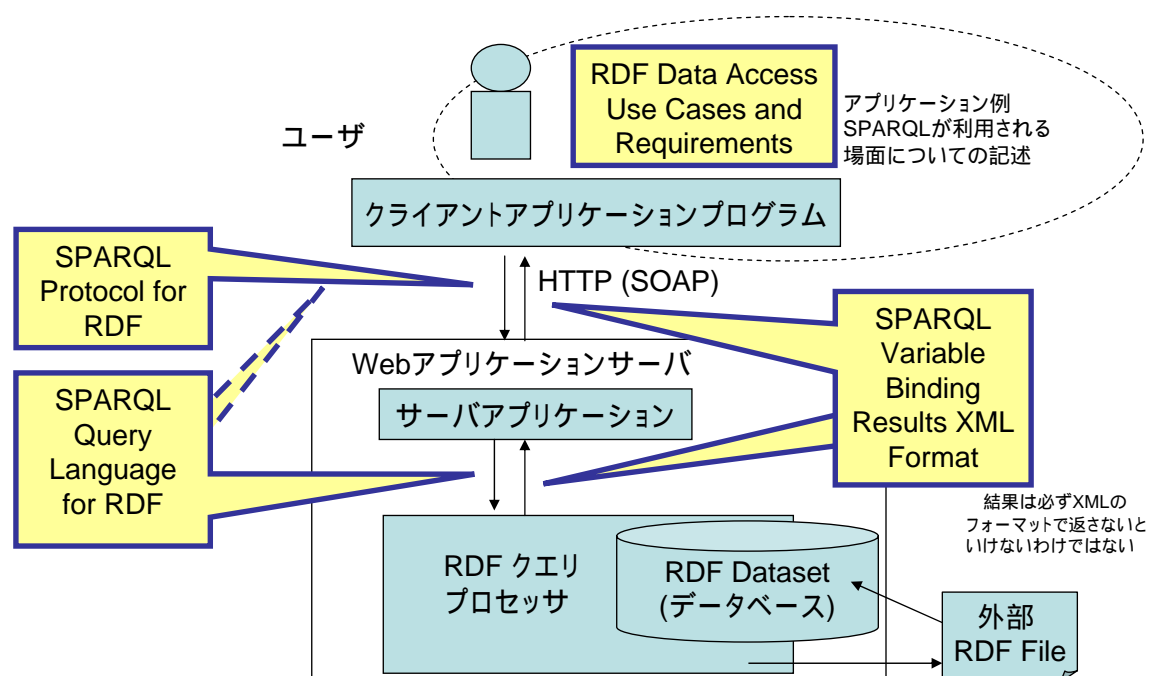
DAWGで策定中の仕様

(現時点ではワーキングドラフトの段階)

- RDF Data Access Use Cases and Requirements
 - W3C Working Draft 12 October 2004
 - <http://www.w3.org/TR/rdf-dawg-uc/>
 - ユースケースと仕様に要求される事項を整理したドキュメント, 仕様策定の範囲などをあらかじめ規定
- SPARQL Query Language for RDF
 - W3C Working Draft 12 October 2004
 - <http://www.w3.org/TR/rdf-sparql-query/>
 - RDFのクエリ言語仕様を規定
- SPARQL Variable Binding Results XML Format
 - W3C Working Draft 21 December 2004
 - <http://www.w3.org/TR/rdf-sparql-XMLres/>
 - クエリの結果の変数と値の組をXMLで返す場合のフォーマットを規定
- SPARQL Protocol for RDF
 - W3C Working Draft 14 January 2005
 - <http://www.w3.org/TR/rdf-sparql-protocol/>
 - クエリプロセッサへのアクセスについて抽象プロトコルとHTTPをベースとしたプロトコルを規定

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

DAWG仕様の位置付け(適用イメージ)



(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

ユースケース

- DAWGでは最初にユースケースを検討し、その後それらを参照しながら仕様のrequirementの議論が行われた
- 各ユースケースにはそれを実現するのに必要と考えられるrequirementへのポイントも示されている
- 以下の場面で必要となるデータアクセスを想定したものが用意された
 - Personal Information Management
 - Supply Chain Management
 - Publishing
 - Multimedia
 - Transportation
 - Tourism
 - Software Development
 - Instructional Technology
 - Social Network Analysis
 - Health Care
 - Market Research
 - Data Aggregation
 - Device Independence

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

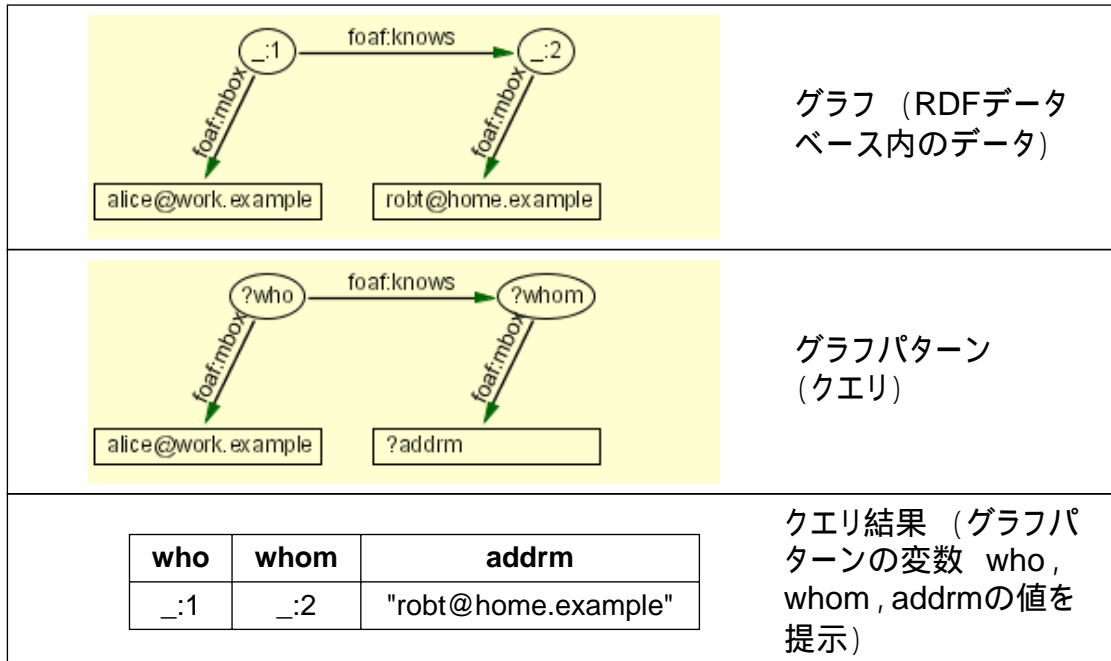
ユースケースの例

- Finding an Email Address (Personal Information Management)
 - メールクライアントソフトから直接RDF (FOAFデータ) で記述されたアドレス帳にクエリを発行してメールアドレスを調べる
- Finding Information About Motorcycle Parts (Supply Chain Management)
 - バイクの1つのパーツに関する情報の問い合わせから関連するパーツの情報を探せる
- Finding Unknown Media Objects (Publishing)
 - さまざまなメディアの巨大複合知識ベースに対して、新しい情報がないか定期的にRDFクエリーを発することで、条件にマッチした情報をe-mailで受け取ったりできる
- Monitoring News Events (Multimedia)
- Avoiding Traffic Jams (Transportation)
 - 道路状況、工事計画、天候などの複数のRDFデータベースに対して車からアクセスして渋滞を避けることができる
- Discovering What People Say about News Stories (Publishing)
- Exploring the Neighborhood (Tourism)
- Finding Out New Things About People (Social Network Analysis)
- ...

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

SPARQL

- SPARQL Protocol and RDF Query Language
- グラフベースのクエリ言語(トリプルパターンマッチング)



(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

SPARQLクエリ仕様(複数マッチング)

- SELECT節
 - 値を取得したい変数の名前を記述
- WHERE節
 - 変数を含んだグラフパターン(トリプルパターン)

データ

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
_:a foaf:name "Johnny Lee Outlaw" .
_:a foaf:mbox <mailto:jlow@example.com> .
_:b foaf:name "Peter Goodguy" .
_:b foaf:mbox <mailto:peter@example.org> .
```

クエリ

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name, ?mbox
WHERE
( ?x foaf:name ?name )
( ?x foaf:mbox ?mbox )
```

変数

トリプル
パターン

クエリ結果

name	mbox
"Johnny Lee Outlaw"	<mailto:jlow@example.com>
"Peter Goodguy"	<mailto:peter@example.org>

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

SPARQLクエリ仕様(値の制約)

データ

```
@prefix dc:
<http://purl.org/dc/elements/1.1/> .
@prefix :
<http://example.org/book/> .
@prefix ns:
<http://example.org/ns#> .

:book1 dc:title "SPARQL Tutorial" .
:book1 ns:price 42 .
:book2 dc:title "The Semantic Web" .
:book2 ns:price 23 .
```

クエリ

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX ns: <http://example.org/ns#>
SELECT ?title ?price
WHERE ( ?x dc:title ?title )
      ( ?x ns:price ?price ) AND ?price < 30
```

値の制約
(priceの値
が30未満の
パターンのみ
マッチ)

クエリ結果

title	price
"The Semantic Web"	23

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

SPARQLクエリ仕様(オプションルマッチング)

データ

```
@prefix foaf:
<http://xmlns.com/foaf/0.1/> .
@prefix rdf:
<http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:
<http://www.w3.org/2000/01/rdf-schema#> .

_:a rdf:type foaf:Person .
_:a foaf:name "Alice" .
_:a foaf:mbox <mailto:alice@work.example> .

_:b rdf:type foaf:Person .
_:b foaf:name "Bob" .
```

クエリ

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE ( ?x foaf:name ?name )
      OPTIONAL ( ?x foaf:mbox ?mbox )
```

オプションルパターン
(このパターンが
マッチしなくても結
果は得られる)

クエリ結果

name	mbox
"Alice"	<mailto:alice@example.com>
"Bob"	

[(?x foaf:mbox ?mbox)]
と記述することが可能

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

今後の課題(1) (現在WGで検討中の仕様)

- Nested Optional Blocks
 - ネスト構造になったオプショントリプルパターンを含んだグラフパターンによるクエリを可能にするか？
- ORやNOTに相当するマッチングを可能にするか？
 - UNION
 - 代替(alternative)パターンのマッチング
 - disjunction
 - UNSAID
 - 節に記述したトリプルパターンがマッチしないグラフから解を得る
 - Negation as Failure

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

今後の課題(2) (現在WGで検討中の仕様)

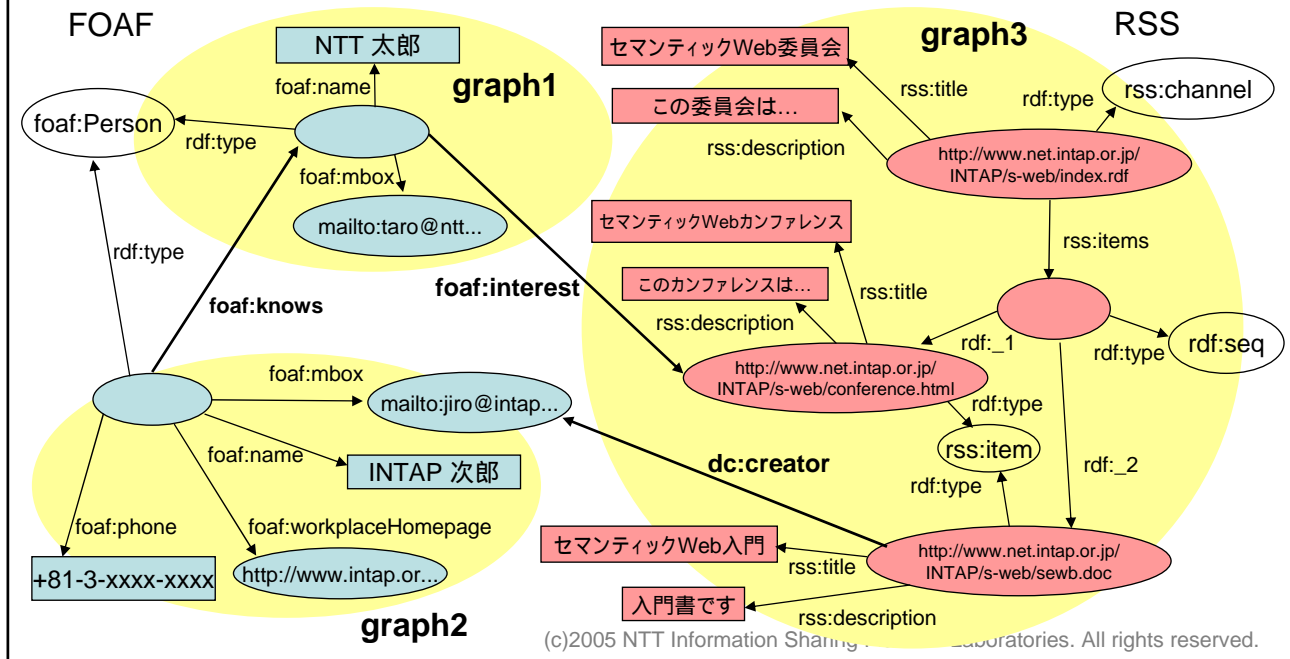
- クエリ対象の指定
 - クエリプロセッサが扱うRDFデータ(データセット)内のグラフに対してURIを与えるなどしてクエリ対象を明示
 - named graph
- クエリ結果がどのグラフから得られたか(SOURCE)
 - 複数のグラフを含むデータセットに対するクエリから得られた結果がどのグラフから得られたかも提示
 - また、クエリ対象を指定したクエリ範囲の制約記述を可能にする
- クエリ結果の返し方
 - 変数と値の組以外の結果提示方法
 - テンプレートを利用して結果の値を含んだサブグラフを返す(CONSTRUCT)
 - 結果を含むグラフを返す(DESCRIBE)
 - マッチするトリプルパターンがあるかないかだけをYes/Noで返す(ASK)

これらの仕様のいくつかは今後の検討において削除される可能性があります

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

検討中の仕様が実装されると...

- 複数のWebサイトに分散して存在するFOAFデータ(**graph1**および**graph2**)とRSSデータ(**graph3**)を利用して、以下の情報やその出所を簡易に取得できるようになる
 - 「INTAP 次郎」の知人が興味を持っているページの要約
 - 「セマンティックWeb入門」の著者の電話番号や働いている場所



WGの今後の予定

- 前述のRequirementsにあって、まだ確定していない仕様の検討を継続
- テストケースの検討を継続
 - <http://www.w3.org/2001/sw/DataAccess/tests/>
- 2005年3月中旬には仕様のLast Call Working Draftを発行予定
 - 当初予定では2005年7月にW3C勧告

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.

おわりに

- 既に実験的に仕様の実装が行われている
 - DawgShowsではWebフォームにRDFデータとクエリを入力すると結果を返すデモを公開
 - <http://esw.w3.org/topic/DawgShows>
 - SPARQLer - An RDF Query Demo
 - Andy Seaborne (HP)
 - Redland Rasqal RDF Query Demonstration
 - Dave Beckett (Bristol大)
 - (SPARQL検討のベースとなったクエリ言語であるRDQLを実装したミドルウェアは複数存在する)
- 仕様に対するコメントは以下のアドレスで受け付けている
 - public-rdf-dawg-comments@w3.org

本資料の内容は2005年1月時点の情報に基づいたものです。今後仕様などは変更される可能性があります。

本資料の作成に関してW3CのEric Prud'hommeaux氏、松下電器産業の福重貴雄氏から有益なコメントをいただきました。記して感謝の意を表します。

(c)2005 NTT Information Sharing Platform Laboratories. All rights reserved.