

文脈情報を利用した、リンクデータの表現の変換規則と変換方法 についての一考察

Representations and Transformations of Linked Data Using Context

水野隆文^{1*}

Takafumi Mizuno¹

¹ 名城大学

¹ Meijo University

Abstract: In usage of linked data, such as RDF, when users obtain data once, then the data are used on multiple occasions. The data are reformulated into any data structures in users' local environments. The reused data are referred to as context in this study. Context is used to generalize and to concrete information.

1 はじめに

一般に、リソース間の関係はリンクにより表現される。RDF (Resource Description Framework) は、Web上に公開されたデータの関係を表示する枠組みである [1]。RDF では、それぞれの関係を、主語 *Subject*、述語 *Predicate*、目的語 *Object* をつぎの三つ組で表現する。

$$\langle \textit{Subject}, \textit{Predicate}, \textit{Object} \rangle \quad (1)$$

主語、述語、目的語は、機械的に判読し継続的に参照をたどっていくことができるように、URI として記述される。主語と述語については空白 (ブランク) を許容する。述語にはさらにリテラルの記述が可能である。これら RDF の三つ組を要素とする集合を本稿では RDF データとよぶ。RDF データを格納し、必要とする者やサービスにこれらを提供するシステムを RDF ストアという。Web 上には、数十の単位から、数千万単位の三つ組を格納する RDF ストアが存在する。RDF を利用する場合は、RDF ストアに問い合わせを行い、必要な RDF データを抽出する。

RDF データは、機械の可読性を持たせるかわりに、人間の可読性を下げている。また、URI の文字列を大量に

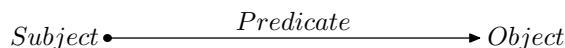


図 1: RDF の三つ組は、述語 *Subject* を始点、目的語 *Object* を終点、述語 *Predicate* を辺のラベルとした有向グラフとして表現できる。

格納した RDF ストアから Web を介して検索を行うため、問い合わせによっては高コストとなる。そのため、取得した RDF データをローカル環境や使用目的に合わせて、表現を変換して保存することがある。

Spivak は、RDF データと、(スキーマの) 圏から **Sets** への関手として表現されるデータモデルとの変換方法を提示した [2]。

関手的データモデル まず、Spivak が提案した関手的データモデル [2] を説明する。 $G = (V, A)$ を有向グラフとする。 V は頂点 (オブジェクト) の集合、 A は矢 (射) の集合である。矢 $a \in A$ は、頂点 $src(p)$ から出て頂点 $tgt(p)$ に入る。このグラフのパスの集合を $Path_G$ とする。有向グラフ G と、パスの集合 $Path_G$ におけるある合同関係 \simeq の組 $S := (G, \simeq)$ はスキーマ (圏) である。この合同関係 \simeq はつぎの条件を満たす。

1. $p \simeq q, p, q \in Path_G$ ならば、 $src(p) = src(q)$ 。
2. $p \simeq q, p, q \in Path_G$ ならば、 $tgt(p) = tgt(q)$ 。
3. $p, p', q, q' \in Path_G, a = src(p) = src(p'), b = tgt(p) = tgt(p'), b = src(q) = src(q'), c = tgt(q) = tgt(q')$, そして $++$ を 2 つのパスを連結する演算子とする。このとき、 $p \simeq p'$ かつ $q \simeq q'$ ならば、 $(p ++ q) \simeq (p' ++ q')$ 。

このスキーマをもつデータ (インスタンス) J はつぎのように S から集合の圏 **Sets** への関手により定義される。

$$J : S \rightarrow \mathbf{Sets}. \quad (2)$$

データは随時更新されており、そのときそのときの状態を表現するのが関手 J である。関手的データモデルに

*連絡先: 名城大学
〒461-8534 愛知県名古屋市東区矢田南 4-102-9
E-mail: tmizuno@meijo-u.ac.jp

においては、更新のようなデータの状態の変化は関手間の自然変換として表現される。

関手的データモデルと RDF との変換 圏 S と関手 J から構成される圏 $\int_S J$ を考える。 $\int_S J$ の対象の集合は

$$\{(X, x) | X \in S, x \in J(X)\}, \quad (3)$$

対象 (X, x) から (Y, y) への射の集合は

$$\{f : X \rightarrow Y | J(X)(x) = y\} \quad (4)$$

とする。 $\int_S J$ におけるある対象 (X, x) とその対象から別の対象 (Y, y) への一つの射 $h : (X, x) \rightarrow (Y, y)$ が RDF の三つ組 $\langle x, h, y \rangle$ に対応する。 $\int_S J$ から S へはつぎの射影関手 π_J により変換される。

$$\pi_J : \int_S J \rightarrow S \quad (5)$$

π_J は自然に定義され、 $\pi_J((X, x)) = X$, $\pi_J(f)((X, x)) = f(X)$ である。

2 文脈

実際の RDF の利用では、RDF ストアへ同じ問い合わせを何度も行うのは高コストとなる場合がある。プログラミングにおいては、一度取得した RDF は目的に合わせたデータ構造に変換し記憶する。変換されたデータは、元の RDF の構造に戻ることができる。上記の記号では、変換されたデータは、ある関手 I を用いて $I(S)$ と表現される圏の対象である。この対象がリンクデータの表現であり、関手 I や π_J がその変換規則である。

一般に RDF はメタデータを表現しており、RDF の三つ組に記述できるリテラルには制限がある。三つ組には基本的に、Web 上に存在するデータへの参照が記述されており、そのデータの実体は別のデータベースに格納されている。

RDF の利用に際しては、ローカルの環境から Web にアクセスし、まず RDF ストアからメタデータやグラフデータ、あまり大きくないテキストデータ、他の RDF ストア、外部データベースに存在するデータへの参照を取得する。RDF から必要なデータを取得した後は、別の RDF ストアや別のデータベースに接続し、さらにデータを取得していく。このようにデータをつぎつぎに取得していくときには、一度取得したデータはローカルの環境に保存し、それを使い回していくことが行われる。このローカルの環境に記憶されるデータは、Web を介したアクセスを繰り返すたびに単調に増加していく。本稿では、このようにローカルに保存されるデータを文脈とよぶ。図 2 はアクセスを繰り返すたびに文脈が C' , C'' , C''' と変化していく様子を表現している。

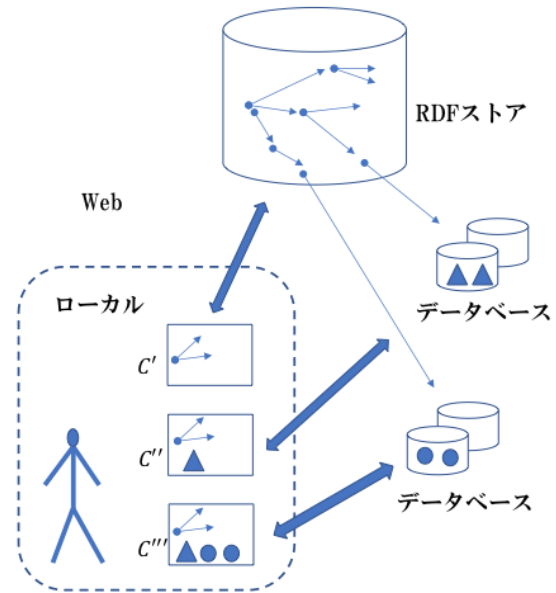


図 2: 一度 Web から取得した情報をローカルの環境で加工して使い回す。

3 文脈の利用

ローカルの環境に蓄積される RDF データは、利用の目的、再検索のコストなど、RDF のネットワークで表現されていない知識を用いて変換されることがある。このようなときには文脈に依存し変換されると考えられる。(5) の変換では、 S の構造が先にあることが前提であった。複数の RDF の三つ組の主語 x_1, x_2, \dots をまとめる概念 X がある場合に $(X, x_1), (X, x_2), \dots$ から π_J により変換できる。プログラミングにおいては、文脈の変化に応じて、 x_i や $I(x_i)$ を代表する型やデータ構造のスロット、データを記録するキャッシュファイルなどを用意し名前をつけることに相当する。これにより、代数的データとして RDF の情報を扱うことができる。

アプリケーションの設計の視点からみた場合、文脈は、情報の抽象化と情報の具体化に利用される。情報の抽象化では、2つのデータのグラフ表現について、共有する部分グラフ以外の情報を削除し変数に置き換える。具体化では、モデルに文脈の情報を追加しておくことにより、モデルにさらに追加できる情報を制限する。

参考文献

- [1] Hayes, P.: RDF Semantics, W3C Recommendation (2004) <http://www.w3.org/TR/2004/REC-rdf-nt-20040210/>
- [2] Spivak, D.: *Category Theory for the Sciences*, MIT Press.