

Linked Data の構造分析に基づく高速道路立ち寄り推薦サービス Recommending Stops along Highway based on Structural Analysis of Linked Data

大河原 渉^{*1} 牧山 宅矢^{*1} 大野 祐^{*2} 森田 武史^{*3} 杉山 岳弘^{*2} 山口 高平^{*1}
Wataru Okawara Takuya Makiyama Yu Ono Takeshi Morita Takehiro Sugiyama Takahira Yamaguchi

^{*1} 慶応義塾大学
Keio University

^{*2} 静岡大学
Shizuoka University

^{*3} 青山学院大学
Aoyama Gakuin University

This paper discusses how to produce and consume Linked Data to develop recommendation service for stopping to Highway users. We automatically translate relevant web pages into Linked Data by scrapers and produce Linked Data across eight domains in order to execute recommendation facilities, including spot recommendation with user context. The field test with 60 or more Highway users has been done at two Service Areas in Shizuoka West Region. It has shown us which recommendation facilities go well or not. We discuss recommendation services with Linked Data with the filled test results.

1. はじめに

Linked Open Data(LOD)は,RDF と呼ばれる最小単位でデータを扱い,さまざまなドメインをまたいでデータをつなぎ合わせる事(拡張容易性)によりデータ利用者の利便性を向上させるプラットフォームとして取り組みが進められている。一方で,Linked Data の有用な使い方については確立されておらず,明確な効用評価もされていないのが現状である。

本稿では,高速道路立ち寄りというタスクに着目し,Web ページからスクレイピングを行い Linked Data を作成する。さらに現在公開されている Linked Open Data を組み合わせることによって 8 つの異なるドメインを有する Linked Data を構築する。ユーザコンテキストを Linked Data として抽出し,作成した Linked Data のドメインをクロスすることによりユーザに適合性の高い立ち寄りスポットを推薦するシステムを提案する。

2. 高速道路立ち寄り推薦サービス

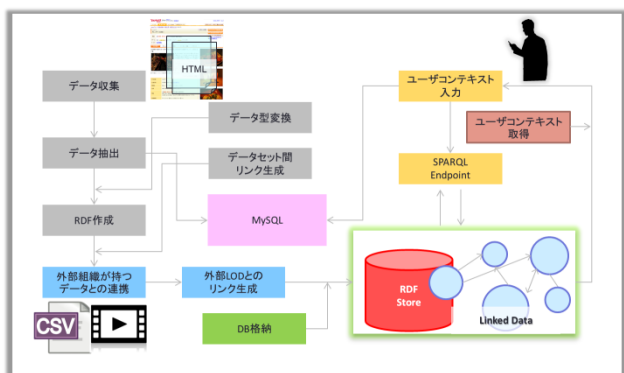


図 1 システムモジュール構成図

2.1 概要

システムは大きく「Linked Data 生成」と「アプリケーション」に分かれる。Linked Data 生成では Web ページからデータを収集しデータ型を変換して RDF を作成する。さらに,既存の LOD と組み合わせることによって Linked Data を構築する。アプリケーション

ン部分でユーザコンテキストに基づき Linked Data にクエリを投げかけるための SPARQL エンドポイントを作成する。

2.2 Linked Data の生成

2.2.1 データ収集及びデータ抽出

現在,観光情報や高速道路の日本語で作成された Linked Data は存在しない。そこで Web サイトから情報を収集する。

さらに HTML ファイルから半自動的に必要な部分を抽出する。HTML の構造は Web ページごとに異なるためスクレイピングプログラムを作成した。正規表現,DOM TREE 構造を利用した方法を用いデータの抽出を行う。

2.2.2 RDF 生成

抽出したデータに対してモデリングを行い,「RAP(RDF API for PHP)」¹と呼ばれる PHP ライブラリを用いて,RDF データへの変換を行う。また,述語を表現する際には標準語彙(Linked Open Vocabularies)²及び独自に定義した語彙を用いて記述する。

さらに,所在地や営業時間,料金などの情報はリテラル値になっているためコンピュータが自動的に処理することができない。そこでリテラル値を機械可読なデータ型に変換する必要がある。図 2 にデータ型変換後の RDF モデル例を示す。

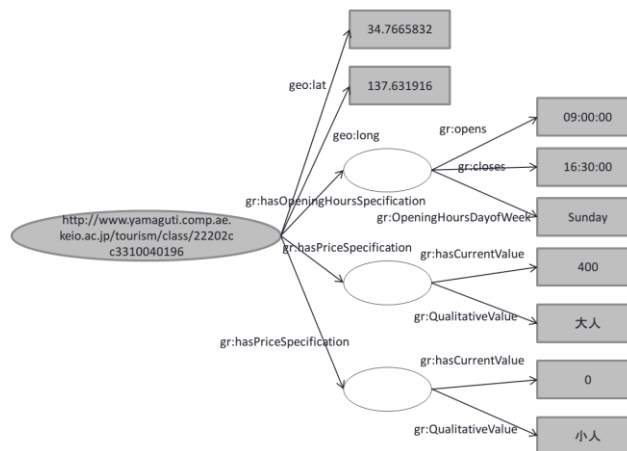


図 2 データ型変換後の RDF モデル

連絡先：牧山宅矢, 山口高平
慶應義塾大学理工学部管理工学科
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
TEL:045-566-1614
E-mail: shioman0@gmail.com,yamaguti@ae.keio.ac.jp

¹ RAP: <http://wifo5-03.informatik.uni-mannheim.de/bizer/rdfapi/>

² Linked Open Vocabularies: <http://lov.okfn.org/>

2.2.3 データセット間リンク生成

RDF トリプルはデータセットごとに生成されているため、この時点ではデータセット間に関係性がない。そのため、データセット間にリンクを生成する必要がある。図 2 は観光情報と高速道路情報に関するリソースにリンクを生成した例である。

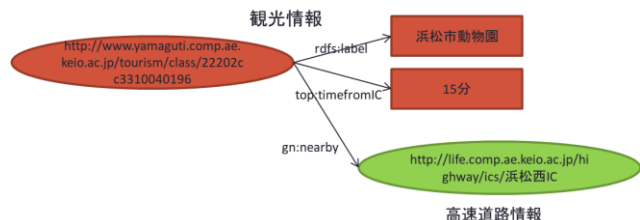


図 3 gn:nearby を用いたリンク生成例

観光情報のデータセット観光スポットは緯度経度の位置情報を持つ。また、高速道路情報の IC についても同様に緯度経度を持つ。そこで位置情報を用いてそれぞれのスポットの最寄りの IC を調べるにより、「gn:nearby」という周辺にあるという意味を持つ標準語彙を利用したプロパティを用いることによってデータセット間にリンクを生成している。

2.2.4 外部 LOD との連携

現在公開されている LOD と作成した Linked Data との連携を行う。本稿では、外部リンクとしてふじのくにオープンデータカタログ¹と日本語 Wikipedia オントロジー[玉川 11]を利用した。ふじのくにオープンデータカタログは防災情報や統計情報など静岡県が所有するさまざまな行政データをオープン化したものである。外部 LOD とのリンクを生成する際、原則としてまずリソースの文字列の完全照合を行うことにより、両者のリソースが同義であるかどうかを判断する。しかし、文字列の表記ゆれを含んでいるものが多いためリンク生成の精度が低くなってしまいます。また、本稿で扱う LOD は位置情報(緯度経度情報)を含んでいるため以下の判断基準でリソースの一致を行う。

(1) スポットの文字列の部分照合を行う

表記ゆれに対応するため、それぞれの名称から空白や「・」を取り除いた上で、文字列を比較、照合するか判断する。

(2) それぞれの地点間の距離が一定基準以内である

それぞれの緯度経度から求めた直線距離が一定基準以内(100メートル以内)であることを条件とする。

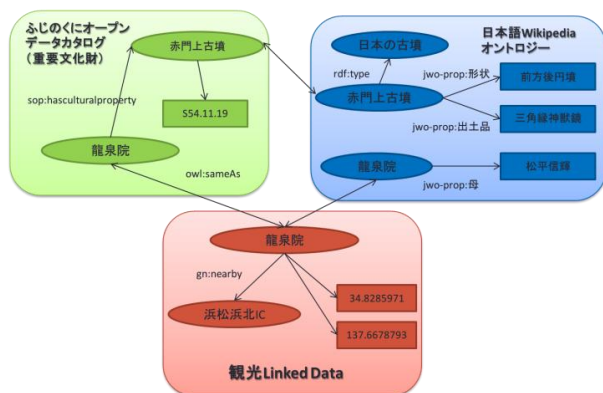


図 4 LOD リンク生成の例

図 4 のリンク生成例では観光情報に関する Linked Data とふじのくにオープンデータカタログ、日本語 Wikipedia オントロジーの 3 つのデータセット間にリンクを生成している。観光情報の Linked Data には龍泉院というスポットが存在するがスポットの詳細情報に関するトリプルが存在しない。そこで、ふじのくにオープンデータカタログにある龍泉院のリソースにリンクを生成することにより、龍泉院にある文化財の情報を取得することができる。さらに、日本語 Wikipedia オントロジーにリンクを生成することにより文化財の詳細情報やクラス-インスタンス関係を取得することが可能になる。また、観光情報の Linked Data と日本語 Wikipedia オントロジーにリンクを生成することにより、縁のある人物など日本語 Wikipedia オントロジーに特有の情報も取得することが可能である。

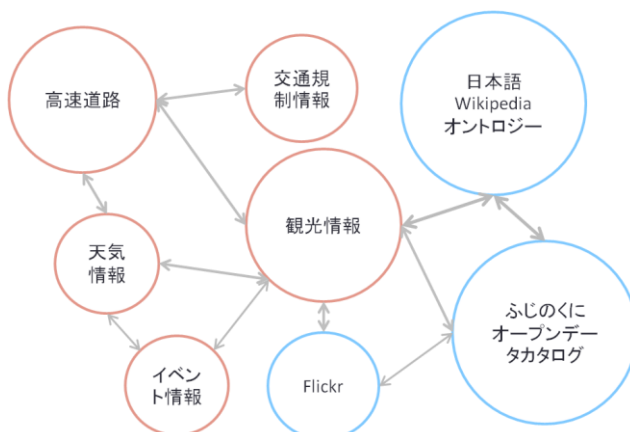


図 5 構築した Linked Data

2.2.5 構築した Linked Data

今回作成した Linked Data を図 5 に示す。赤い丸の部分は今回作成したデータセットであり、青い丸の部分は既に存在する LOD で計 8 つの異なるデータセットから構成されている。総トリプル数は 169097 トリプルである。以下では今回作成した Linked Data のハブとなっている観光情報についてのデータセット内の RDF モデル例を示す。

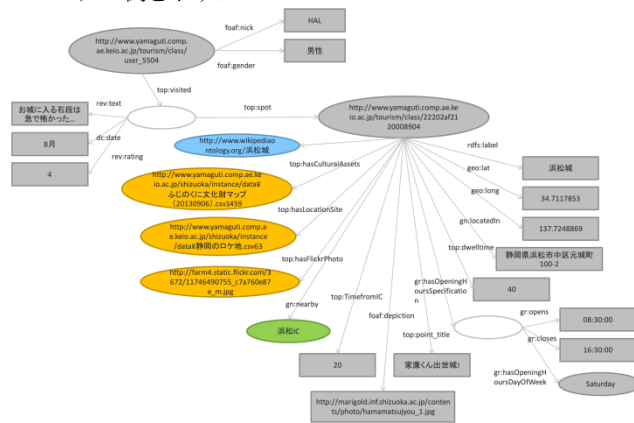


図 6 観光情報に関する RDF モデル

観光スポットの RDF モデルはスポット名、緯度経度、住所などの基本情報、口コミ情報から構成されている。さらに、日本語 Wikipedia オントロジーやふじのくにオープンデータカタログの文化財、パワースポット、ロケ地に関する情報、Flickr、高速道路情報などの外部のデータセットにリンクされている。

¹ ふじのくにオープンデータカタログ : <http://open-data.pref.shizuoka.jp>

2.2.6 Linked Data 格納

今まで作成してきた Linked Data を MySQL データベースに格納する。この際に「ARC2 RDF CLASSES FOR PHP¹」と呼ばれる PHP ライブラリを利用した。一方で、全てのデータに対して同様の処理を行うことは現実的ではない。その原因として、ストリーミングデータが挙げられる。これらのデータは非常に速いサイクルで最新データに変わり、過去のデータは必要とされなくなってしまうからだ。そこで、本稿では、スタティックデータのみデータベースに格納して、ストリーミングデータに関しては、都度 API を利用して RDF データを取得する仕様とした。

2.3 アプリケーション

2.3.1 ユーザコンテキスト抽出

作成した Linked Data へ SPARQL クエリを投げるためのユーザコンテキストを抽出する。年齢、性別、目的地についてはアプリケーションの最初に入力を行う。位置情報についてはユーザが利用を許可することにより取得される。予算、立ち寄りの目的、時間に関してはアプリケーション内で入力を行う。

一方で表 1 に記載した情報については Linked Data からコンテキストを取得する。

表 1 Linked Data で取得したコンテキスト一覧

交通規制情報	方向	上り/下り
	規制内容	渋滞中 / 通行止 / 特になし
気象情報	天気	〇〇℃
	気温	晴れ/曇り/雨/雪

交通規制情報に関しては、はじめに入力した目的地と自動的に取得した現在地の緯度経度を比較することにより道路名及び方向を取得し、道路と方向に一致する規制情報を取得する。図 7 にユーザ設定を行いそこから交通規制情報を取得する流れを示す。気象情報に関しては、まず位置情報から最寄りの IC を取得する。さらに図 5 からわかるように IC と天気情報はリンクが張られているためそこから IC 周辺の天気情報を取得する。

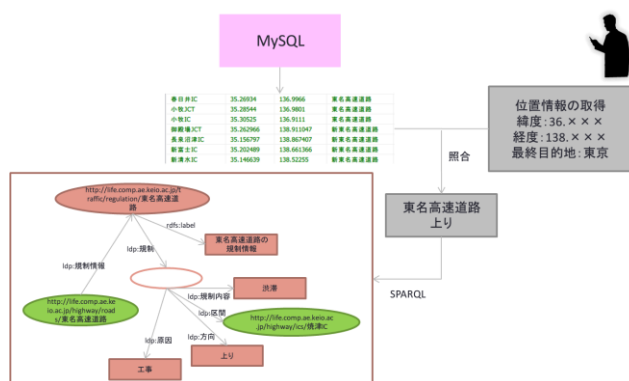


図 7 ユーザ設定からの交通規制情報取得

2.3.2 SPARQL Endpoint

作成した Linked Data に対して検索できるように「SPARQL Endpoint」を作成した。そこに目的、ユーザセグメント、予算、観光時間、現在時刻など抽出したユーザコンテキストを用いて検索を行うことでスポット情報を取得する。現在時刻に関してはスポットの営業時間と現在時刻を比較することで営業時間内伸びスポット

を推薦することが可能になる。このように SPARQL を用いることでさまざまなユーザコンテキストを用いた柔軟なクエリに対応することができる。

2.3.3 スケジューリング

スケジューリングではユーザの位置情報、観光情報にあるスポットの位置情報、交通規制情報のドメインをリンクすることにより、現在地から観光スポット、観光スポットにおける滞在時間、観光スポットから最終目的地までの時間を表示することによりユーザの立ち寄り行動の時空間管理を行う。

2.3.4 機能一覧

ここでは今回構築した Linked Data を利用したアプリケーションの機能一覧を示し、Linked Data を利用したアプリケーション構築の事例として紹介する。本稿では高速道路利用者を対象に、高速道路を走行中及び高速道路上の SA/PA で休憩をしているときに利用することを想定しアプリケーションを開発した。また、移動中の利用も考えられるため、スマートフォン上で動作する仕様とした。以下で実装した機能について述べる。

(1) ユーザコンテキストに基づくスポット推薦

ユーザとのインタラクションを通してユーザコンテキストを取得し、ユーザコンテキストから該当するスポットを抽出し推薦する機能である。ユーザとの簡単な質問形式のインタラクションを通して、旅行の目的として食べる/見る・遊ぶ・体験・癒し、グループ構成として家族/カップル、夫婦/男性グループ/男女グループ/女性グループ、スポットで使える予算、時間を取得する。このようにして取得したユーザコンテキストを利用し、Linked Data に対してクエリを投げることによりユーザにマッチしたスポットを取得する。本機能では Linked Data を利用することで現在の位置情報、方向から進行方向にある IC 周辺の、現在時刻と営業時間を比較することにより現在営業中のスポットなど、さまざまな条件に当てはまるスポットを1つのクエリを通して抽出することが可能となる。

(2) Linked Data を利用したスポット検索

ユーザコンテキストとは関係なく Linked Data を利用してさまざまなスポット情報を表示し、ユーザが自分で探索を行う機能である。ここでは検索の中の1つの例であるにロケ地情報の検索例について図 8 に示す。



図 8 ロケ地情報に関する検索

ロケ地情報に関しては、まずふじのくにオープンデータからスポットに関するロケ地情報として作品名と簡潔な説明を取得する。さらに日本語 Wikipedia オントロジーを利用することで作品名から作品のキャストや詳細情報を閲覧することが可能になっている。

(3) 開催中のイベント情報

¹ ARC2 : <https://github.com/semsol/arc2>

更新の頻度が高い情報について Linked Data 作成し、それを利用して実装した開催中のイベント情報についての機能を説明する。開催中のイベント情報に関しては静岡新聞 SBS の Web サイトを利用し、本日開催されているイベント情報に関してスクレイピングを行い、情報を抽出した。イベント情報は一日ごとに更新されるものもあれば数週間更新されないものもある。そこで、一日ごとに Linked Data を作成し、MySQL に追加することにした。イベント情報に関してはイベントの基本情報や Linked Data でドメインをたどることによって開催されている地域の天気情報、会場になっているスポットの基本情報、口コミ情報を閲覧することができる。

3. 高速道路立ち寄り小規模社会実験と評価

高速道路からの立ち寄りは、観光地や自宅など本来の目的地に行くという予定を変更し、さらに高速道路料金を余計に払わなければならない行動であり、コストや時間の面から考えると大変難しいテーマである。本稿では、Linked Data を利用し、さまざまな情報やコンテンツを統合することで本来の予定を変更する立ち寄り行動を促すことが目的である。そこで評価実験では、ユーザに実際にアプリケーションを利用してもらい、さらにアンケートに回答してもらうことによりユーザにとって有用な Linked Data の組み合わせ方をシナリオベースで考察する。さらにユーザのページアクセスのログを記録することによりアプリケーションの評価を行う。表 2 に今回行った小規模社会実験の内容を示す。

表 2 実験内容

	浜松 SA(上り)	浜名湖 SA
調査日	平成 26 年 1 月下旬	
1 回目アンケート方法	調査員による聞き取り調査	
2 回目アンケート方法	Web アンケートによる回答	
1 回目回収枚数	39 枚	24 枚
2 回目回収枚数	3 枚	1 枚

今回の実験では計 63 名の被験者のうち 4 人が立ち寄りを行った。さらに、立ち寄り可能性のある被験者として未定の 12 名のデータを抽出し、立ち寄りを行った人がアプリケーションをどのような経緯で利用し立ち寄りに至ったかを Linked Data のクロスドメイン、コンテンツ、検索方法の 3 つの評価要素から考察する。

(1) クロスドメイン

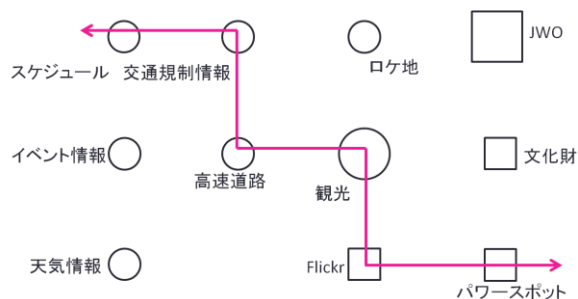


図 9 クロスドメインの例

以下にシナリオ例を示す。カップルでサファリパークに行った帰り、SA で休憩した。アプリケーションを使うとパワースポットについて検索できるので利用する。いくつかの候補の中からフラワーパークについてのスポット情報を見る。Flickr の写真を見て、夜のイルミネーションも綺麗であることがわかり立ち寄りしたいと思う。さらに、帰りの時間が気になるので、スケジュール機能を利用して最終目的地である名古屋に到着する時間を確認する。

図 9 から、シナリオ例では観光、高速道路、交通規制情報、スケジュール、Flickr、パワースポットの計 6 つのドメインをクロスしてい

ることがわかる。また Flickr やパワースポットなどオープンデータにもクロスしており、立ち寄り行動に Linked Data のクロスドメインが有用であると言える。

(2) コンテンツ

コンテンツについては、立ち寄りを行った半数以上のユーザが写真を決め手にしている。さらに IC からの時間や観光時間を気にするユーザが多い。一方で、動画や口コミを参考に立ち寄りユーザはいなかった。

(3) 選択スポット数

表 3 検索方法の分類

選択スポット数	検索方法
多い	スポット一覧、イベント一覧、無料スポット
普通	浜松名物、ロケ地、文化財めぐり、パワースポット、動画で選ぶ
少ない	あなたにオススメ、静岡大学学生オススメ

表 3 のように、選択スポット数に応じて検索方法を分類し評価を行った。ユーザのページアクセスのログをとると、立ち寄りを行ったユーザの半数以上が、選択スポット数が少ない検索方法を選び、ある程度絞られた中で選択すること(推薦)を好むことがわかった。

立ち寄りを行うユーザは普通、その土地についての情報はあまり知らない。そこで、一覧からスポットのリストを見て自分で探索するよりも静岡大学学生オススメの機能などを用いることによって地元の人が提供する信頼性のある情報を参考にすることがわかった。

実際に立ち寄ったユーザの特徴としては比較的時間に余裕があるということが挙げられる。一方でそのようなユーザでも営業時間を確認するために役に立った機能としてスケジューリング機能を上げており、当初は料金という要素が立ち寄りにおいて重要であると考えられていたが、時間が重要な factor であるということがわかった。このことは、コンテンツの部分で動画はあまり役に立たず、軽量のマルチメディアである写真が効果的であるということにも反映されている。

4. おわりに

本稿では、位置情報を持つ観光情報や高速道路情報に着目して Linked Data を作成し、また現在公開されている Linked Open Data を組み合わせることによって 8 つのドメインを有する Linked Data を構築した。さらに、それを利用して高速道路利用者の立ち寄り推薦サービスを開発した。評価を行うために小規模の社会実験を実施し、移動中における立ち寄り行動の分析を行うとともに、Linked Data のクロスドメインに着目して、レコメンデーションシステムにおける Linked Data の有用性を示すことができた。

謝辞

本研究は、中日本高速道路株式会社からの支援を受けて実施されており、ここに記して感謝します。

参考文献

[玉川 11] 玉川奨, 森田武史, 山口高平: 日本語 Wikipedia からプロパティを備えたオントロジーの構築, 人工知能学会論文誌 26(4), 504-517, (2011)
 [保科 12] 保科宗淳, 大河原渉, 山口高平: 領域オントロジーと Linked Data を利用した観光情報推薦, 第 26 回人工知能学会全国大会, 3C2-OS-13b-8, (2012)