

地域メディア情報を活用した Linked Data サービスの検討と課題

A Study on Linked Data Services for Regional News Media

長野 伸一^{1*} 川村 隆浩¹ 小林 巖生²
Shinichi Nagano¹ Takahiro Kawamura¹ Iwao Kobayashi²

¹ (株) 東芝 研究開発センター

¹ Corporate R&D Center, Toshiba Corporation

² NPO 法人リンクト・オープン・データ・イニシアティブ

² Linked Open Data Initiative

Abstract: Regional news media has a potential of being the hub which connects people, stuff, and information in the region. This paper proposes Linked Data enhanced services for the media, and then discusses the issue of building Linked Data from textual data of the news.

1 はじめに

Web が我々の生活や社会に浸透するにつれて、社会や産業の基盤となりうるデータは、人々や組織が単独で保有するよりも、互いにオープンにして共有化するほうが、所有者にとっても社会にとってもデータの価値が高まるとの考え方が広がっている。特に、2013年6月、閣議決定された「世界最先端 IT 国家創造宣言」や、G8 サミットで採択された「オープンデータ憲章」において、オープンデータ推進が大きく取り上げられたことをうけ、行政が保有する公共データの民間開放への取り組みが各国で加速しつつある。オープンデータは、行政と民間との連携による行政サービスの改善や、民間による新たなサービス創造が期待されている一方、オープンデータ活用に関する市民参加型のワークショップが各地で開催されるなど、市民レベルでオープンデータを推進する取り組みも進められている。さらに、行政が保有していない地域情報や生活情報を市民の手で創ることで、地域の活性化や様々な産業分野での活用への期待も高まっている。

本稿では、来るオープンデータ社会に向けて、情報基盤としての地域メディアの新しい可能性について述べる。地域のオープンデータを連携させるハブとして地方新聞の可能性に注目し、メディア情報の Linked Open Data 化、および地域コミュニティの活性化に向けたサービス提案を行う。

2 地域メディアの可能性

我々市民は、日々の生活の中で、様々な地域情報に触れている。例えば、住まい、子育て、教育、介護など暮らしの情報や、地域店舗の特売やイベント開催などの生活密着型の情報が含まれる。また、道路、環境、防災などまちづくりに関するものや、地域を訪れる観光客向けの情報も含まれるであろう。インターネットや Web、携帯電話の普及により、我々は、様々な情報源から多様な情報チャンネルを通して、これらの地域情報を受信するとともに、我々自身も情報を発信している。

近年、地域メディアは、こうした情報流通の担い手の1つとして着目されている。コミュニティFMは代表的な地域メディアの1つとして、地域住民が必要とする、きめ細かな地域情報の伝達手段として各地で定着しつつある。東日本大震災の発生時においては、地域性と即時性を備えたラジオが活用されたことは記憶に新しい [1]。一方で、ラジオは、音声配信であるために情報を再利用しづらく、また双方向性に欠ける。

コミュニティFMと同様に、地域情報の伝達手段として根ざしたメディアとして、地方新聞がある。新聞は文字を中心としたメディアであるため情報の再利用性が高く、地域のヒト、モノ、情報をつなぐ情報基盤として、次の可能性を有している。

- 情報の地産地消

地元の記者や市民が取材した地域の出来事を地元から発信するとともに、情報がソーシャルメディア等を通して地域に拡散され、地元で消費される。

- コミュニティづくり

*連絡先: (株) 東芝 研究開発センター 知識メディアラボラトリー
〒 212-8582 神奈川県川崎市幸区小向東芝町 1
E-mail: shinichi3.nagano@toshiba.co.jp

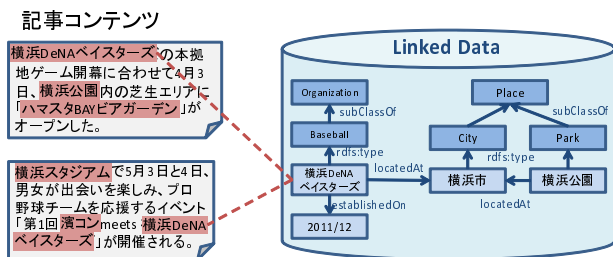


図 1: 記事コンテンツの Linked Data 化

地域情報が話題の種となって、共通の関心を持つ仲間が生まれ、住民に新たな行動を喚起するきっかけとなる。

● 情報連結による価値向上

地方新聞の有する情報が、地域のオープンデータ等と連携することにより、情報が地域内でつながり価値が高まる。

3 地域メディア情報活用サービス

地域メディアの例として、ヨコハマ経済新聞¹を取り上げる。ヨコハマ経済新聞は、全国に約 90 あるオンラインの地域新聞サイト「みんなの経済新聞ネットワーク」²の1つである。2004年にWebサイトがスタートし、2013年7月現在、約 8000 件の記事が公開されている。主に、横浜で開催されるイベントや、商業施設の開店などが取り上げられている。以降では、これらの情報をまとめてイベントと呼ぶ。現在のWebサイトでは、新着順の閲覧、およびキーワード検索が、記事への主な動線である。

記事の内容が表すイベント情報は、その属性(人、組織、場所、日時、テーマ)により特徴づけられる。本稿では、記事コンテンツのテキストから、属性情報を抽出し、Linked Data化を試みる。記事コンテンツのLinked Data化のイメージを図1に示す。Linked Data化することにより、ノードを介して記事どうしに明示的な関連性を付与できる。

Linked Data化によりコンテンツ資源としての再利用性が高まるため、従来のWebサイトに対して、新たな付加価値を提供するサービスの実現が可能となる。地域メディア情報を活用したサービスとして、以下の4つを提案する。

1. 記事のまとめ

人、組織、テーマをトピックとして、関連する複数の記事をまとめたページを作成し、記事とペー

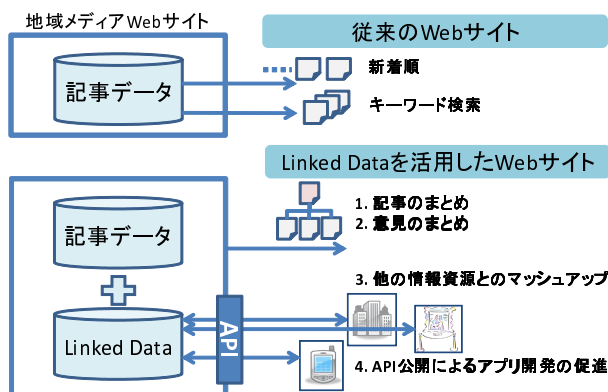


図 2: 地域メディア情報活用サービス

ジの間にリンクを付与する。利用者が記事のリンクを順次たどることによる、Webサイト上で旅するような新たなユーザ体験を提供する。

2. 意見のまとめ

上記の記事まとめに加えて、各トピックに関連するソーシャルメディア情報を収集することにより、地域住民や観光客によるクチコミを提示し、利用者に行動を喚起する。

3. 他の情報資源とのマッシュアップ

自治体が保有するオープンデータやイベント情報と連結することで、地域情報に対して、多角的な視点を与える。

4. API公開によるアプリ開発の促進

API公開によるデータ提供を通して、他事業者によるサービス開発や、ハッカソン等のイベントによる市民アプリ開発を促進する。

地域メディアを Linked Data化するメリットとして、第1にサイトへのPV向上があげられる。記事や意見をまとめた新たなページを作成することによるトラフィックの増加に加えて、記事コンテンツ間のリンクにより、サイト滞在時間の増加が見込まれる。第2に、Linked Data化したデータセットのAPIを公開することにより、他事業者や市民によるデータの再利用を促進し、これまでに見られなかった発想によるマッシュアップサービス、アプリからの新たなトラフィックが創出される。第3に、記事コンテンツの句の情報、自治体・市民が持つ地域資源と連結することで、地域情報の再構成、再発見が促進され、コンテンツ価値の向上が期待される。

4 Linked Data構築

ヨコハマ経済新聞のWebサイトから入手した幾つかの記事コンテンツを題材として、記事の内容が表す主

¹<http://www.hamakei.com/>

²<http://minkei.net/>

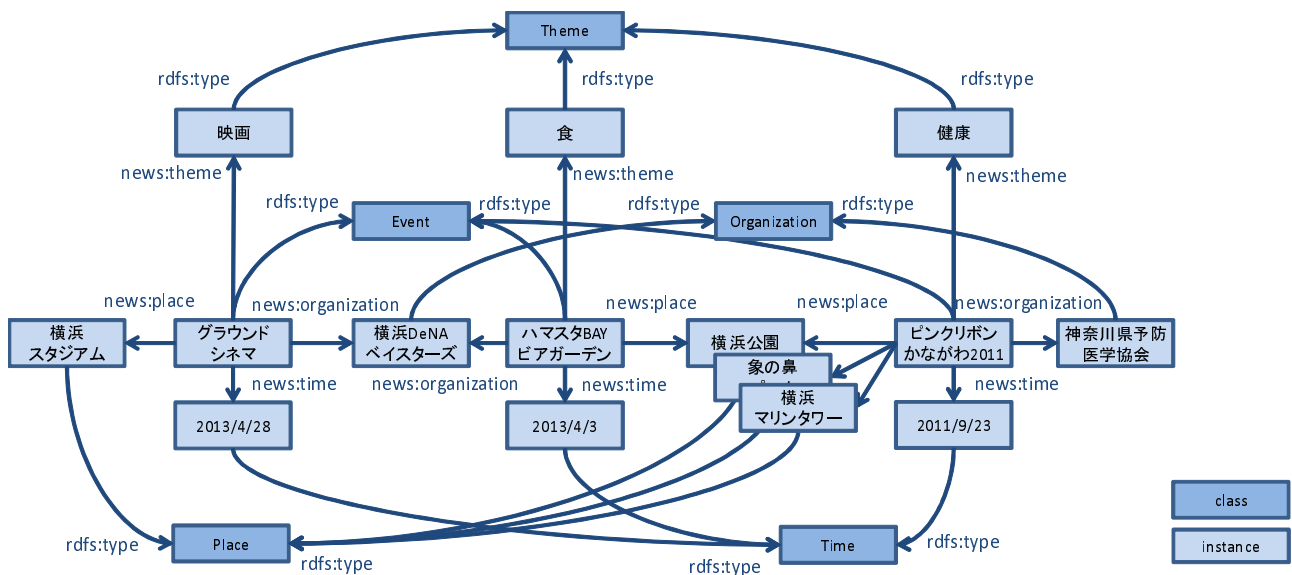


図 3: Linked Data 構築の例

題に関する属性情報(人, 組織, 場所, 日時, テーマ)を Linked Data 化する際の課題について述べる。

各記事コンテンツは, 記事 ID, 見出し, 本文から構成されていると仮定する。見出し, 本文ともプレーンテキストであり, メタデータは付与されない。また記事の一部に含まれる画像データ, 記事中に出現する他記事へのリンク, 過去の関連記事へのリンクは利用しない。記事コンテンツの例を表 1 に示す。属性の多くは見出しと本文 1 文目とに出現するほか, 表現に幾つかのパターンが存在する(例えば「<場所>で<日付>, <イベント>が開催される」)。

表 1: 記事コンテンツの例

ID	http://www.hamakei.com/headline/7820/
見出し	横浜公園に「ハマスタ BAY ビアガーデン」 ー大型ビジョンで野球中継も
本文	横浜 DeNA ベイスターズの本拠地ゲーム開幕に合わせて 4 月 3 日、横浜公園（横浜市中区横浜公園）内の芝生エリアに「ハマスタ BAY ビアガーデン」がオープンした。昨年のゴールデンウィークと 8 月の計 22 日間に実施したビアガーデン企画を、今年は拡大して...(略)

地域メディア情報の Linked Data 化は, 見出しおよび本文に対して自然言語処理を適用し, 記事が表すイベントの属性(人, 組織, 場所, 日時, テーマ)を抽出することにより行う。地域メディア情報からの Linked Data 構築には, Corenews Ontology[2] を利用する。Corenews Ontology は英 BBC が開発したオントロジーで, ニュー

ス記事のメタデータ記述に利用される。ニュース記事の人, 組織, 場所, テーマ, イベントの 5 つの要素で構造化し, RDFa を利用して Web ページへメタデータを埋め込んでいる。簡潔な構造で取り扱いやすいため, 今回試作に利用した。

複数の記事コンテンツから人手でイベントと属性とを抽出し, Linked Data 化したグラフの例を図 3 に示す。3 件の記事³それぞれから抽出した計 3 件のイベント「グラウンドシネマ」「ハマスタ BAY ビアガーデン」「ピンクリボンかながわ 2011」を Linked Data 化したものである。「グラウンドシネマ」と「ハマスタ BAY ビアガーデン」はいずれも主催者が「横浜 DeNA ベイスターズ」であり, 「ハマスタ BAY ビアガーデン」と「ピンクリボンかながわ 2011」はいずれも「横浜公園」で開催されている。このように, Linked Data 化した属性情報を利用してイベントどうしを連結できる。

5 考察

地域メディア情報の Linked Data 化における課題に関して, 本試作を通して明らかになった課題を定性的に述べる。

まず, 自然言語処理に関して述べる。第 1 に, 固有表現抽出の難しさがある。「平成 16 年度第 1 回横浜観光プロモーションフォーラム認定事業」や「名取市図書館どんぐり子ども図書室」など, 名称が非常に長いものが存在する。こうした名称は括弧付きで表記されるも

³<http://www.hamakei.com/headline/7872/>
<http://www.hamakei.com/headline/7820/>
<http://www.hamakei.com/headline/6402/>

のも多いが、インタビューのコメントや、記事中の強調でも括弧が利用されており、区別が必要である。第2に、表記揺れによる曖昧性がある。見出しと本文とで、あるいは初出しとそれ以降とで表記が異なる場合がある。例えば、ある記事の見出しでは「濱コン×横浜 DeNA ベイスターズ」とあるが、記事本文では「第1回濱コン meets 横浜 DeNA ベイスターズ」と表記されている。第3に、名称のないイベントが存在する。例えば、「現在使用されていない歴史的建造物を文化・芸術で活用し都心部活性化を図るプロジェクト」は、どこまでが名称を表すか人間が読んで解釈が難しい。第4に、一部のイベント間に、全体-部分の関係がある。例えば、ビデオアート展「彷徨（ほうこう）～都市と砂漠の間で」は、「横浜フランス月間2013」の一環で開催されている。こうした関係を高い精度で抽出することにより、Linked Dataの密なグラフを構築できる。

次に、Linked Data構築に関して述べる。第1に、グラフの断片化が予想される。地元住民や観光客が多く訪れる著名な施設やイベントでない場合、記事への出現回数が少なく、抽出した属性情報から構成したグラフが他のノードとの間にエッジを持たず、孤立した部分グラフが作成される。記事中で言及されるエピソードなどを活用して、記事どうしを連結させる工夫が必要である。第2の課題は、再利用のしやすさである。記事では組織やイベントの名称は正式なものが利用される傾向にある。APIとして公開する場合、目的によっては愛称や略称で検索できると使い勝手がよい。第3は、精度とコストの問題である。テキスト解析による自動処理だけでは精度に限界がある。CMSを活用して記事執筆を支援するツールを整備するなど、人手による構築や編集と併用した運用プロセスを定める必要がある。

6 関連研究

マスメディアのデータをLinked Data化した最初の事例の1つは、英BBCの取り組みである。BBCは、テレビやラジオの番組情報をLinked Dataとして公開し、従来のWebページと組み合わせたサイトを構築した[3]。また、サッカーWorld Cup 2010やロンドンオリンピックのBBCサイトの構築では、選手やチーム、試合などの情報をRDFでモデル化することにより、BBC内部でのデータ統合が容易となったという。また、米New York Times⁴は、記事の主要な件名標目をLinked Dataとして公開している。件名標目をフィードとして流通させることで、彼らが保有する豊富なコンテンツ資源へのトラフィックを高めることがねらいであ

る。また、NHK⁵や日本テレビ⁶がテレビ番組メタデータにアクセス可能なAPIを公開したり、丹南ケーブルテレビ⁷が番組で紹介したお店の情報をオープンデータとして公開するなど、日本でもメディア情報の二次利用が広がり始めている。

7 まとめ

地域のオープンデータを連携させるハブとして地方新聞の可能性を考察し、メディア情報を活用したLinked Dataサービスの提案、およびその実現に向けた課題について述べた。今後は、プロトタイプを試作を通して、サービスの有用性検証、および精度評価を進めていく。

なお、その他本論文に掲載の商品、機能等の名称は、それぞれ各社が商標として使用している場合がある。

謝辞

本稿をまとめるにあたり、貴重なご意見をいただきました。ヨコハマ経済新聞編集長 杉浦裕樹氏に心より感謝申し上げます。

参考文献

- [1] 総務省: 災害時における情報通信の在り方に関する調査結果 (2012)
- [2] Jeremy Tarling: Corenews Ontology, <http://www.bbc.co.uk/ontologies/news/2013-05-01.shtml> (2013)
- [3] Georgi Kobilarov, et. al: Media meets semantic web - how the bbc uses dbpedia and linked data to make connections, Proc. of ESWC2009, pp.723-737, 2009.

⁴<http://data.nytimes.com/>

⁵<http://www2.nhk.or.jp/api/>

⁶<http://www.ntv.co.jp/appli/api/index.html>

⁷<http://miseban.com/opendata/>