

## 地域メディア情報を活用した Linked Data サービスの試作評価

## A Prototype of Linked Data Services for Regional News Media

長野 伸一\*1      川村 隆浩\*1      小林 巖生\*2      杉浦 裕樹\*3  
 Shinichi Nagano      Takahiro Kawamura      Iwao Kobayashi      Hiroki Sugiura

\*1(株) 東芝 研究開発センター  
 Corporate R&D Center, Toshiba Corporation

\*2NPO 法人 リンクト・オープン・データ・イニシアティブ  
 Linked Open Data Initiative

\*3NPO 法人 横浜コミュニティデザイン・ラボ  
 Yokohama Community Design Lab.

Regional news media has a potential of being the hub which connects people, stuff, and information living in the region. The paper presents publishing the media contents in the Linked Data manner and its prototype service.

## 1. はじめに

本稿では、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着目し、地域メディアが有するコンテンツ資源の Linked Data 化、およびその活用サービスの試作について述べる。

## 2. 情報基盤としての地域メディア

地域メディアは、地域社会における情報流通の担い手の1つとして着目されている。文字や映像を中心とした情報は再利用性が高く、地域のヒト、モノ、情報をつなぐ基盤としての存在意義は大きい。地域メディアが有する特性として、(1) 情報の地産地消、(2) 地域コミュニティの活性化、(3) 他のオープンデータとの連結による価値向上、などが考えられる [1]。

ヨコハマ経済新聞\*1は、横浜都心臨海部のビジネス&カルチャーニュースを配信する、オンラインの地域新聞である。2004年の開設以来、8500件を超える記事を公開している。主に、横浜都心臨海部で開催される芸術文化や市民主催のイベントや、商業施設の開店などが取り上げられており、地域メディアとして広く浸透している。記事の内容が表す主題は、その属性(人物、組織、場所、製品、イベント)により特徴づけられると考えられる。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。Linked Data化することにより、コンテンツ資源として再利用性が高まり、新たな付加価値の提供が可能となる。

## 3. 記事コンテンツの Linked Data 化

ヨコハマ経済新聞が公開している約8500件の記事コンテンツを題材として、記事の内容が表す主題に関する属性情報を Linked Data 化する。各記事コンテンツは、記事 ID、見出し、本文、公開日の4つの要素から構成されている。見出し、本文ともテキスト情報であり、メタデータは付与されていない。本文は、全文ではなく、第一段落のテキスト情報のみを対象とす

	組織	イベント
記事ID	http://www.hamakei.com/headline/7820/	
見出し	横浜公園に「ハマスタBAYアガーデン」	1階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アガーデン」がオープンした。	
場所	横浜公園	

図 1: 記事データの例

る。また、本来の記事コンテンツの一部に含まれる画像データ、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。

地域メディア情報の Linked Data 化は、見出および本文に対して、人手による編集作業により、記事の内容が表す主題に関する属性(事象)を抽出することにより行う。事象としては、人物、組織、場所、製品、イベントの5つを抽出する。

- **人物**: 人物の名前。読み仮名やニックネームも人名とする。創作上の人物は含まない。例えば、林市長、中畑監督、なかはたきよし、ハマの大魔神が該当する。
- **組織**: 複数の人間からなる組織の名前。企業、団体、法人、チーム、劇団等など。横浜 DeNA ベイスターズ、横浜市、日産自動車、東急電鉄、崎陽軒などが該当する。
- **場所**: 場所に関する名前。市区町村、地域、河川など。横浜、みなとみらい 21 地区、日本大通りなどが該当する。
- **製品**: 人間が製造したもの、創造したものの名前。創作上の人物、芸術作品、商品を含む。ドラえもん、キャッツ、タイタニック、クイーンエリザベス号などが該当する。
- **イベント**: 出来事、催し物に関する名前。映画やミュージカルなどの創作物の名称は含まない。横浜マラソン大会、ヨコハマトリエンナーレ 2014 などが該当する。

Linked Data 記述には、Schema.org, DCMI Metadata Terms などの語彙を利用する。使用したプロパティは以下のとおりである。

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

\*1 <http://www.hamakei.com/>

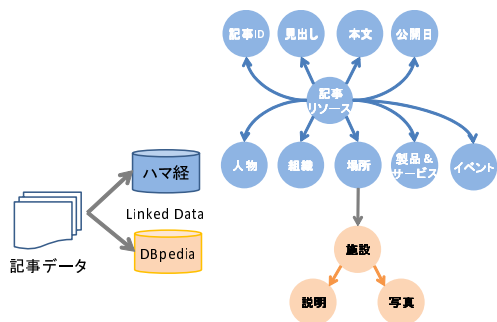


図 2: データモデル

● 記事リソース

語彙	説明
rdf:type	リソースタイプ (schema:NewsArticle)
dcterms:identifier	記事 ID
schema:headline	見出し
schema:text	本文
schema:datePublished	公開日
schema:mentions	事象リソース

● 事象リソース

語彙	説明
rdf:type	リソースタイプ (schema:Person, schema:Organization, schema:Place, schema:Product, schema:Event のいずれか)
rdfs:label	事象の名称
schema:name	事象の名称
owl:sameAs	外部リソースへの参照

4. 考察

地域メディア情報の Linked Data 構築に関して、[1] で示した課題をベースに、本試作で明らかになった課題を述べる。

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。また、記事と DBpedia とで名称が異なるものが存在する。例えば、記事では「赤レンガ倉庫」と表記されるのに対して、DBpedia では「横浜赤レンガ倉庫」と表記され完全一致しない。このように、名称が異なるリテラルどうしのマッピングする仕組みが必要である。

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も解釈が難しい。



図 3: 試作サービスの画面

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」は、「横浜フランス月間 2013」の一環で開催されている。こうした事象リソース間の関係を抽出することにより、記事リソース間の関係性が蜜なグラフを構築できる。

5. ハマ経クロニクル

ハマ経クロニクルは、Linked Data 化したヨコハマ経済新聞の記事コンテンツを活用した、試作サービスである。2013 年に配信された約 1000 件の記事コンテンツを対象にして、横浜都心臨海部に位置する横浜赤レンガ倉庫など 7 つの観光施設それぞれに関する記事を時系列順に閲覧し、1 年間の変化を振り返ることができる。観光施設の名称をキーにして Linked Data のデータセットを検索し、各観光施設が出現する記事の情報を取得している。このように、記事コンテンツを串刺しすることで、最新の情報と過去の情報とをつなぎ、地域情報の再構成・再発見を促進することが容易となる。

6. まとめ

地域メディアが有するコンテンツ資源の Linked Data 化、およびその活用サービスの試作について述べた。今後、Linked Data 構築の自動化について検討を進めていく。tj

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。

参考文献

[1] 長野伸一, 他: 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, セマンティック Web とオントロジー研究会 (2013)

[2] Georgi Kobilarov, et. al: Media meets semantic web - how the bbc uses dbpedia and linked data to make connections, Proc. ESWC2009 (2009).

[3] Evan Sandhaus: Linked Data And The New York Times, Proc. ISWC2009 (2009).