

自治体広報情報の RSS に基づく地域課題分析の試み

A Trial Analysis of Local Problems based on RSS of Public Information by Local Governments

古崎 晃司*¹
Kouji Kozaki

山本 泰智*²
Yasunori Yamamoto

*¹ 大阪大学産業科学研究所 The Institute of Scientific and Industrial Research (ISIR), Osaka University
*² ライフサイエンス統合データベースセンター Database Center for Life Science, ROIS

It is an important issue to solve local problems because they are related to civic life directly. However, it is very difficult to get the total picture of variety of local problems. In order to overcome the difficulty this research aim to develop an information platform which can be used or analysis of local problems based on public information by local governments. The authors focus on RSS published at official websites by local governments. This paper introduces a Linked Open Data (LOD) developed by collecting RSS a year form 80 local governments. We also discuss a trial analysis of the LOD using its SPARQL endpoint.

1. はじめに

誰でも自由に利用可能なオープンデータの公開・活用により、様々な社会問題の解決や新産業の創出を目指す取り組みは世界的に注目されており、2013年 G8 サミットでは「オープンデータ憲章」が採択されるなど [外務省 13a, 13b], 世界各国でその推進が行われている。国内においては、2011年から開催されている「Linked Open Data チャレンジ」[乙守 15]をはじめとしたオープンデータ活用に関するコンテストや、「ハッカソン」と呼ばれる市民参加型の技術開発イベント等も多く行われている。また国内の行政機関においても、2012年7月の「電子行政オープンデータ戦略」の策定をはじめとしたオープンデータ推進の取り組み[浅野 15]や、ICT 技術を活用した地方創生に向けた取り組みが進められている。

このように様々な取り組みにおいてオープンデータの活用は、地域問題を情報技術で解決しようとする Civic Tech と呼ばれる新しい行政と市民の関わりから注目されている。前述のコンテストやハッカソンにおいても、地域課題の解決を目指した取り組みが多く見られ、Code For と呼ばれる各地域に在住する IT 技術を持った市民が集まり、自分たちの住む地域をよくするための活動を進める団体も日本各地で生まれている。

しかし地域課題は市民生活に直結しているが故に、その地域の地理的特性、文化的・歴史的背景、産業構造、住民の特性など、様々な要因が関係しており極めて多様性が高く、その全体像を把握することは極めて難しい。この問題を解決すべく、本研究では地域課題を横断的に分析できる情報基盤を開発することを目指す。

その実現のためのアプローチとして、各自治体が定期的に公開している広報情報に注目した。なぜなら自治体の広報情報は、各自治体が市民に伝えたいと考えている情報であるため、それぞれの地域特性や地域課題が反映されていると思われるからである。そこで本研究では、自治体の広報情報を地域横断的に収集することで、地域課題を横断的に分析できる情報基盤を開発することを目的とする。

自治体が公開している広報情報には、広報誌、Web サイト、SNS など様々なメディアが用いられているが、本研究では自治

体の Web サイトの最新情報を利用する。自治体の Web サイトは広報誌のような紙媒体のもの比べて更新頻度が高く、SNS などに比べて配信される情報に偏りが少ないと思われる。また、自治体の Web サイトの多くは、最新情報を RSS(RDF Site Summary/Really Simple Syndication / Rich Site Summary)¹で配信しており、異なる自治体のサイトであっても統一的な手法で収集できるという利点がある。

このような背景のもと、本論文では、自治体の Web サイトから試験的に収集した RSS を対象とした地域課題の分析を試行する。具体的には、約 80 の地方自治体の Web サイトから 1 年間にわたり収集した RSS を、Linked Open Data(LOD)として公開したもの[古崎 16]を用いて、地域課題の分析に役立つ情報が得られるかを試験的に分析した試みについて報告する。

以下、2 章では自治体広報情報の RSS を収集するシステムについて述べ、3 章では約 80 の自治体を対象に 1 年間に渡って収集した広報情報の LOD 化について述べる。続く 4 章では LOD 化した広報情報の試験的な分析事例について述べ、5 章で本論文を総括するとともに今後の課題について議論する。

2. 自治体広報情報 RSS の収集

ここでは、本研究で用いる自治体広報情報の RSS を収集するシステムである PUSH 大阪&PUSH 広報について述べる。2.1 節では、その開発経緯とシステムの概要について述べ、2.2 節では RSS 収集機能の実装の詳細を述べる。

2.1 PUSH 大阪&PUSH 広報

PUSH 大阪²は、2014 年 8 月に大阪で開催されたハッカソン「Civic Hack Osaka 2014」において、筆者らが行政職員と一般市民の合同チームで開発した、市政情報発信アプリケーションである。このシステムはハッカソン終了後も開発が続けられ、大阪市主催の「大阪から考える CivicTech アプリコンテスト」³においては、アプリ・Web サービス部門のグランプリ(最優秀賞)を受賞している(2105 年 3 月)。

開発の動機は、同グループとなった自治体職員から出された

連絡先: 古崎晃司, 大阪大学産業科学研究所 知識科学研究分野, 〒567-0047 大阪府茨木市美穂ヶ丘 8-1, TEL:06-6879-8416, kozaki@ei.sanken.osaka-u.ac.jp

¹Web サイトの最新情報を配信する仕様には、RSS1.0, RSS2.0, ATOM など複数ある。本論文では、それらすべてを収集対象とする。本文ではそれらを代表して RSS という表記を用いる。

² <http://push.jp.net>

³ <http://www.city.osaka.lg.jp/shimin/page/0000289410.html>

- 「行政が発信する情報が、なかなか市民に届かない」
- しかし、新しい仕組みを導入する際に、「行政職員の作業コストを増加させたくない」

という課題を解決することにある。

その解決のアプローチとして、

1. すべての情報を一律に提供するのではなく、利用者が関心のある情報のみを配信する
2. 自治体の HP で既に配信されている RSS を利用することで、低コストでサービスを実現する

という手法を用いている。

図 1 に PUSH 大阪のシステム概要を示す。大阪市および大阪府 24 区がホームページで配信している RSS を定期的に収集し、RSS の内容を解析して独自のメタデータを付与した後、RDF データベースに格納する。RDF のデータモデルは、RSS のモデルをそのまま利用するのではなく、SPARQL クエリによるデータアクセスの利便性を考慮した独自の RDF モデルに変換している。

エンドユーザ向けのアプリケーション(図 2)⁴では、ユーザが指定した設定(カテゴリ・地域)に応じて、RDF データベースに格納した情報を SPARQL クエリで取得し、ユーザに提示する。アプリケーションは、Web サービス、および、スマートフォンアプリ(Android 版、iOS 版)として提供している。Android 版では配信された情報を定期的に通知バーに表示することで、ユーザが普段見落としがちな情報に気づかせる工夫がなされている。

開発当初の PUSH 大阪は、大阪市および大阪府 24 区が配信している RSS のみを収集対象としていたが、ユーザからの要望に応じて、地域活動協議会など市民向けの情報が発信している団体のブログや Facebook ページの情報も収集対象に追加した。

さらに、大阪市以外の自治体を対象とした「PUSH 広報」も同様の仕組みで開発し、2015 年 1 月より約 80 の自治体のホームページの RSS を情報収集対象に加えた。その後も、収集対象とする自治体を追加し、2016 年 2 月現在では、約 100 の自治体を対象としている。

2.2 RSS 収集機能の実装

RSS の収集機能は、RSS の各バージョンおよび ATOM などのサイトの更新情報を表す主要なメタデータの処理に対応したオープンソースソフトウェア ROME⁵ を利用し、Java を用いて実装した。

システムへの入力は、

1. 収集する RSS のリスト(自治体名、RSS のタイトル、RSS の URL、カテゴリ情報の組)
2. カテゴリ判定に用いるキーワードリスト
3. キーワード解析に用いるキーワードリスト

である。

RSS のカテゴリは、大阪市が公開している RSS のカテゴリ⁶を参考に、life(暮らし)、child(子育て)、health(健康)、event(イベント・観光関係)、senior(高齢者の方へ)、handicap(障がいのある方へ)、town(まちづくり)、job(しごと情報)、gov(行政情報)、bid(入札情報)、biz(事業者向け)、other(その他)の 12 種類とした。

⁴ 図中の画面例は、Android 版のダウンロードサイト (<https://play.google.com/store/apps/details?id=net.uedayou.android.pushosaka>) より引用

⁵ <http://rometools.github.io/rome/>

⁶ http://www.city.osaka.lg.jp/main/site_policy/0000000148.html , 大阪市の RSS

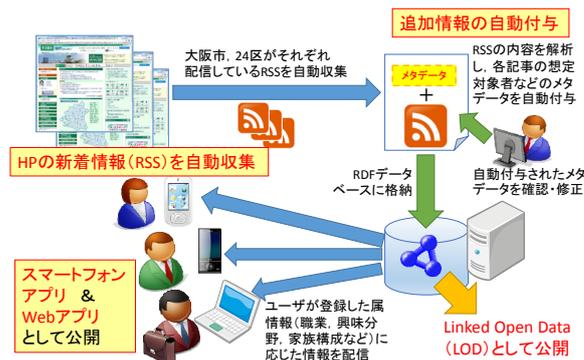


図 1 PUSH 大阪のシステム概要



図 2 PUSH 大阪 (Android 版) の画面例

また、3. のキーワードリストには、大阪市がオープンデータとして公開している統計情報一覧のデータに含まれる検索用キーワード(3,396 単語)⁷を利用している。このキーワードは 50 音順に整理されており、例えば「あ」から始まるものには、「空き地、空き家、悪性新生物、預け金、雨、アメーバ赤痢、アルバイト、アルミニウム、泡現役配送車、安全」が含まれる。

収集する RSS の情報は、いずれの自治体が配信している RSS においても公開されている場合が多い、「各エントリーのタイトル」、「対象とする記事の URL」、「記事の公開日」とした。各エントリーの取得時に、上記 2 および 3 のキーワードリストの中からタイトルに含まれているキーワードを判定し、「カテゴリ指定がされていない RSS に対するカテゴリ判定」および「各記事に含まれているキーワードの取得」を行っている。

収集した RSS の情報は、「自治体名」、「カテゴリ情報」、「キーワード」を付与した後、Turtle 形式で出力し、RDF データベースに格納している。RDF データベースには、Allegro Graph⁸ を利用している。

なお、Facebook ページの情報については、開発当初には提供されていた RSS がサービスの変更に伴い配信されなくなったため、Facebook の GraphAPI を用いて別途、RDF ファイルを生成するプログラムを実装している。

⁷ <http://www.city.osaka.lg.jp/somu/page/0000304815.html#toukei> , 大阪市のオープンデータ (統計情報)

⁸ <http://www.franz.com/agraph/allegrograph/>

3. 自治体広報情報の LOD 化

2015年1月31日から2016年1月28日までの約1年間にわたり、約80の自治体のホームページから収集したRSSファイルを2章で述べたシステムでRDF化した。データの収集は、原則として、平日の午前9時半頃に1日1回行い、各日別ファイルとして保存した。それらのファイルから重複するデータ(対象とするエントリーのURL、公開日時が一致するもの)を除き、データ形式の調整を行った上でTurtle形式のRDFファイルとして保存した。なお、Facebook ページから収集した情報は、利用規約の観点から現時点ではLOD化の対象外としている。

図3にデータモデルおよびデータ例を示す。各リソースのURIは、同じ記事のURLで異なる複数のエントリーで参照される場合(記事の内容が更新された場合など)があることを考慮して、PUSH大阪・広報において独自のIDを付与している。

これらのデータをRDFデータベースに格納し、SPARQLエンドポイントおよびダンプファイルを公開した⁹。公開したデータのサイズは1,667,004トリプルである。

グラフ URI

- ・「PUSH大阪」用(大阪市) : <http://push.jp.net/osaka>
- ・「PUSH広報」の用(大阪市以外) : <http://push.jp.net/koho>

プレフィックス

xsd: <http://www.w3.org/2001/XMLSchema#>
 rss: <http://purl.org/rss/2.0/>
 push: <http://push.jp.net/property#>

プロパティ

push:area RSS 配信元の自治体(英語表記)
 例) "Osaka-city", "Osaka-miyakojima-ku"

push:area_label RSS 配信元の自治体(日本語表記)
 例) "大阪市", "大阪市都島区"

push:category 記事のカテゴリ
 例) "life"

push:category_label 元RSSの名称
 例) "都島区(暮らし)"@ja

rss:title 記事のタイトル (@ja)
 例) "中野地域防災訓練が開催されました"@ja

rss:link 記事のURL
 例) <http://www.city.osaka.lg.jp/miyakojima/page/000240970.html>

push:lodkey 記事のキーワード(複数指定可)
 例) "地域", "防災"

push:lastUpdate RSSの収集日時
 例) "2015-01-09T18:43:09Z"^^xsd:dateTime

rss:pubDate 記事の公開日時
 例) "2015-01-09T18:43:09Z"^^xsd:dateTime

データ例

<http://push.jp.net/resource/data#84170>

```

push:area "Osaka-city";
push:area_label "大阪市";
push:category "other";
push:category_label "大阪市民の方へ"@ja;
rss:title "北区菅原町に歴史顕彰板が完成しました!!"@ja;
rss:link <http://www.city.osaka.lg.jp/toshiseibi/page/0000295402.html>;
push:lodkey "歴史";
push:lastUpdate "2015-01-26T16:12:19Z"^^xsd:dateTime;
rss:pubDate "2015-01-26T16:12:19Z"^^xsd:dateTime.
    
```

図3 LOD化した自治体広報情報RSSのデータモデルおよびデータ例

⁹ http://push.jp.net/data.html

4. 自治体広報情報RSSの分析の試行

3章で述べたLOD化した自治体広報情報RSSの試験的な分析を行った。分析は、SPARQLクエリを用いて様々な条件により検索した結果を利用する。以下、その概要を述べる。

なお紙面の都合で詳細は述べないが、SPARQLの知識がない利用者向けに、GUIによる簡易分析ツール¹⁰も試作している。

4.1 各エントリーに含まれるキーワードの比較

各エントリーのタイトルに含まれるキーワードの数を自治体毎に集計したい場合は、下記のようなSPARQLクエリを用いる¹¹。

```

PREFIX :<http://push.jp.net/property#>
SELECT ?l (count(?l) as ?lc) ?area ?t{
GRAPH ?g{
    ?s :area_label ?l;
        :area ?area;
        :lodkey ?t.
}}group by ?l ?area ?t order by desc(?lc)
    
```

この結果を用いて大阪市、横浜市、室蘭市、金沢市の記事に含まれるキーワード数(上位10位)を比較すると、図4のようになった。これらの結果のうち、各自治体の特徴が表れていると思われるキーワードには、図中の太枠で示している。例えば、同じ大都市であっても大阪では“地域”、“環境”、横浜では“障害”といったキーワードが多く現れている。一方、室蘭市の“空き家”、金沢市の“気象”、“牛”は他の自治体の記事にはあまり含まれていない特徴的なキーワードである。これらの違いは、各自治体の政策や市民が関心を持っている課題の違いを示唆していると思われる。

同様にして、他の自治体の記事に含まれるキーワードのうち、最も出現数が多いものを比較した。その結果、特徴的なものとして、さいたま市・千葉市の“動物”、金沢市の“気象”、糸魚川市の“雪”、神戸市の“美術”、奈良市の“プラネタリウム”、南越前市の“CATV”などが見られた。

これらは単純に、記事に現れるキーワードの出現回数のみを比較しているが、地域ごとの特徴がある程度は見て取れると思われる。これらを他の情報と組み合わせることで、より詳細な分析が可能になると期待される。

大阪市		横浜市		室蘭市		金沢市	
キーワード	件数	キーワード	件数	キーワード	件数	キーワード	件数
施設	166	障害	50	臨時職員	89	気象	167
地域	159	保育所	50	建設	83	牛	118
福祉	154	アルバイト	41	定時	77	科学	55
住宅	149	福祉	38	水道	60	美術	52
介護	91	企業	35	空き家	41	美術館	40
環境	86	介護	30	保育所	31	議会	37
消防	84	大学	30	住宅	29	大学	32
保育所	84	市営	28	議会	22	文化	32
教育	81	地下鉄	24	購入	22	乳幼児	22
法人	80	生活	20	市営	20	感染症	19

※「事業」など一般的なキーワードは除外している。

障害 自治体毎に特徴が現れていると思われるキーワード

図4 記事に含まれるキーワードの比較

¹⁰ http://push.jp.net/dataComp.html

¹¹ グラフURIをクエリに含めているのは、大阪市とその他の自治体のRSSは別々に収集しており、リソースのURIに重複が見られる可能性があるためである。リソースのURIの付与方法については、今後、改善を検討する。

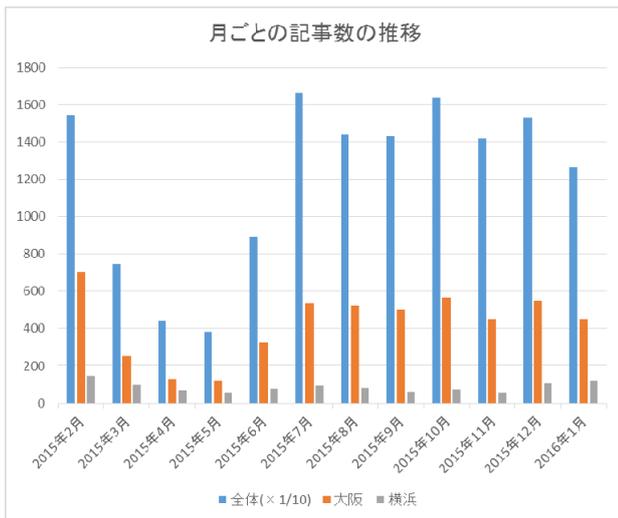


図 5 月ごとの記事数の推移

4.2 時系列による記事数の比較

公開日時毎の記事数を調べる際には、下記のような SPARQL クエリを用いる。

```

PREFIX :<http://push.jp.net/property#>
PREFIX rss: <http://purl.org/rss/2.0/>
SELECT (count(?url) AS ?c) {
?s rss:link ?url;
  rss:pubDate ?date.
FILTER (
?date>="2015-01-01T00:00:00Z"^^xsd:dateTime
&&
?date<"2015-02-01T00:00:00Z"^^xsd:dateTime)}

```

上記の例では 2015 年 1 月に公開された記事数を取得するが、同様のクエリを、対象とする公開日時の区間を 1 ヶ月毎に区切って用いることで、年間を通した月毎の記事数の推移を比較することができる。

これらの推移を、データ全体、大阪市、横浜市の記事数で比較した結果を図 5 に示す。この結果から、全体、両市ともに、年度が明けた後の 4-5 月の記事数が少なく、6-7 月にかけて記事数が急増する傾向が見られる。さらに、月毎の記事に現れるキーワードの数を比較するなど、より詳細な条件を追加することで、様々な観点からの分析が可能であると思われる。

4.3 SNS との組み合わせによる分析

LOD 化した RSS に加え、Facebook や Twitter 等の SNS の情報を組み合わせることにより、情報を発信する自治体側の観点に加え、情報を受け取る市民の観点からの分析が可能になると思われる。

例として、2015 年 8 月 27 日および 30 日に収集した RSS を対象に、各エントリを対象とする URL が、Facebook および Twitter 上で共有された数を調査した¹²。その結果、

- Facebook でのシェアされていた記事の大半は、一部の自治体(大阪市, 大阪市北区, 横浜市)である。
- Twitter でのシェア(ツイート)数は Facebook の2倍以上¹³
- Twitter は犬・猫・動物関係の記事が多く、ツイート数が上位20内の記事の半数を占める

といった傾向がみられた。今後、SPARQL クエリによる分析などとの組み合わせることで、より詳細な分析ができると期待される。

5. おわりに

本論文では、自治体がホームページで公開している広報情報の RSS を収集し、LOD 化したものを対象に、試験的な分析を試みた。対象としたデータは、約 80 の自治体を対象に 1 年にわたって収集したものであるが、地域毎に現れるキーワードの違いや、公開時期による記事数の推移など、地域の特徴を分析するのに有用な情報となり得ることが示唆された。これは、RSS という広く普及した共通フォーマットが、地域横断的なデータの有効活用につながることを示す有益な事例となることも期待される。

今後、より詳細な分析を進めることで提案手法の有用性や課題の考察を進めると共に、収集対象とする自治体を増加させることで、全国横断的な分析を可能とする知識基盤として発展させていく計画である。特に、SNS から得られる情報の受け手である市民の反応や、日本政府や各自治体が公開しているオープンデータから得られる各自治体の特徴に関わるデータを組み合わせることで、より高度な分析を行うことを検討している。例えば、各種統計的な特徴が近い自治体間で、記事に現れるキーワードの傾向を比較することで、似た状況にある自治体の共通課題を分析することなどを検討している。

また、分析に用いるキーワードをオントロジーとして体系化することで、表層的な用語の出現回数にとどまらず、その用語に内在する意味を考慮した解析を検討することも、重要な課題であると認識している。

謝辞

本研究の一部は科学研究費補助金基盤研究(B) 25280081 助成を受けて行った。また、PUSH 大阪の開発に携わった、PUSH 大阪開発 Team のメンバー、特に画面デザインと iOS 版アプリの開発を担当された大西裕子氏、および、Android 版アプリの開発を担当された上田洋氏に感謝いたします。

参考文献

- [外務省 13a] オープンデータ憲章(概要), http://www.mofa.go.jp/mofaj/gaiko/page23_000044.html, 2013.
- [外務省 13a] G8 Open Data Charter, <http://www.mofa.go.jp/mofaj/files/000006820.pdf>, 2013.
- [浅野 15] 浅野優:オープンデータ普及促進に向けた国内行政機関の取り組み, 人工知能学会誌, Vol. 30, No. 5, pp.590-597, 2015.
- [乙守 15] 乙守信行, 中辻真, 萩野達也:オープンデータの普及促進を加速させるコンテストの開催-LOD チャレンジ Japan の取組み-, 人工知能学会誌, Vol. 30, No. 5, pp.598-604, 2015.
- [古崎 16] 古崎晃司, 山本泰智:地域課題分析に向けた自治体広報情報の収集と LOD 化の試み, 人工知能学会研究会資料, SIG-SWO-038-09

¹² 調査結果は、<http://push.jp.net/BBT/fb-ranking.html> (Facebook) および、<http://push.jp.net/BBT/tw-ranking.html> (Twitter) にて公開している。

¹³ Facebook と Twitter の API 仕様などの差による影響も考えられるため、単純に数を比較することはできないと思われる。