

コンテンツ理解をサポートするための オープンデータ利用手法の検討

Study of open data use technique to support the content understanding

宮崎 勝*¹ 浦川 真*¹ 苗村 昌秀*¹ 藤沢 寛*¹
Masaru Miyazaki Makoto Urakawa Masahide Naemura Hiroshi Fujisawa

*¹NHK 放送技術研究所
NHK Science & Technology Research Laboratories

In this paper, we propose a method to present the relevant information for a better understanding of the content of web site. When you select a keyword that is included in the web site while Internet browsing, this prototype service searches for relevant information about the keyword using the various linked open data that has been widely published.

1. はじめに

インターネットの Web サイトの利用は、世の中の情報や知識を得るための一般的な手段となった。日本人の情報行動に関する調査 [総務省 15] によると、「世の中のできごとや動きについて信頼できる情報を得る」際に、インターネットを最も頻りに利用すると答えたユーザの数は、2012 年では全体の 14.7%であり、2000 年時点での 0.4%という数値と比べると、大きく増加している。また、Web の持つ特徴であるハイパーテキストは、情報の補完をするために大きな役割を担っている。例えば、Web サイト閲覧中にハイパーテキストをクリックすると、そのリンク先 Web サイトにおいて、クリックしたテキストに関連した情報を閲覧することができる。このように、Web サイトとそれらをつなぐハイパーテキストにより、ユーザは複数の Web サイトを移動しながら、知りたい情報の詳細や、新たな気づきを得たりすることが可能となった。

一方で近年、インターネットには数多くのオープンデータが公開され、利用されている。例えば、ユーザの集合知を集めた Wikipedia は、物事に関する情報を得る手段として一般的になっており、ユーザの情報探索行動における特徴として、Wikipedia を起点として Web サイトを行き来する行動などが確認されている [安藤 10]。ソフトウェア処理可能な形式で記述されたオープンデータは、Linked Open Data (LOD) と呼ばれる。例えば、日本語版 Wikipedia の情報は、日本語版 DBpedia*¹ において LOD として公開され、利用されている。DBpedia 以外にも、書籍、芸術作品、観光情報などに関する整理・構造化された情報が LOD として公開されており、さまざまなサービスから活用できるようになっている。

本稿では、Web 閲覧中に興味を持ったテーマに関連する情報を、さまざまな LOD を横断的に検索することで容易に得ることのできるシステムを提案する。インターネットブラウザに、ユーザが指定したキーワードに関する情報をさまざまなオープンデータから抽出・閲覧できるインターフェースを提供することで、ユーザの情報探索行動を効果的に支援することが可能となる。

2. Web ブラウザにおける補完情報検索

インターネットブラウザの中には、Web サイト上の興味を持ったキーワードをマウスで選択し、コンテキストメニュー*²を表示させることにより、そのキーワードの解説文などを閲覧できるものがある。例えば、Apple 社の Safari では、Web サイト上で興味を持ったキーワード (例:「鯖江市」) をドラッグして選択し、コンテキストメニューから「鯖江市」について調べる」という項目を選択することにより、「鯖江市」に関する Wikipedia の記事などをブラウザ上にオーバーラップして表示することができる (図 1)。このような機能により、ユーザが閲覧中の Web サイトに関連する情報を、簡単に得ることができる。

筆者らは、ブラウザのこのようなインターフェースを利用した関連情報提示サービスとして、「マイ健康ディクショナリ」を提案している [miyazaki 15]。マイ健康ディクショナリは、Web サイト閲覧中に気になった健康関連キーワードをマウス選択することにより、そのキーワードに関連したテレビ番組の情報を表示する機能を持つ。ユーザはその番組情報を起点に、健康



図 1: Safari における Wikipedia 情報の表示
(出典:総務省ホームページ)

*² GUI で表示されたコンテンツの中のアイテムをクリックすることで表示されるポップアップメニュー

番組の Web サイトを閲覧したり、番組の動画を VOD(Video on Demand) で視聴することが可能となっている。また、「知識マップ」と呼ばれる構造化知識を利用することにより、ユーザが選択したキーワード A に直接関連する健康関連番組の情報を提供するだけでなく、A に関連した別のキーワード B を知識マップから取得し、B に関連した別の番組などを提示する連想検索も可能となっている。図 2 に、マイ健康ディクショナリの画面イメージを示す。図 2 の例では、ユーザが「脳梗塞」というキーワードを選択し、さらにそのキーワードに紐付いた関係である「予防法」を選択した際に、「脳梗塞の予防法」に関連した番組の一覧を表示している。さらに、脳梗塞の予防に効くと言われている「黒酢」のレシピを紹介した料理番組なども表示することができる。

次章では、このマイ健康ディクショナリを拡張して実現した、さまざまなオープンデータを横断的に検索・表示できる“Open Knowledge Dictionary”サービスの概要について述べる。

3. Open Knowledge Dictionary の開発

オープンデータを統合的に扱うサービスとしては、世界中で公開されている LOD を収集して検索することのできる活用基盤、LOD4ALL が提案されている [naseer 14]。従来、特定の対象に関する情報をさまざまな LOD から検索する際には、個々の LOD に対して個別にクエリを発行し、情報を抽出したうえで統合しなければならなかったが、LOD4ALL のような活用基盤を利用することにより、ユーザが興味を持った対象に関する情報を、世界中の LOD のの中から見つけ出すことができる。

一方筆者らは、ブラウザ上に表示されている Web コンテンツに関連した情報を、公開されているさまざまな LOD を横断的に検索することで取得し、ブラウザ上に提示するサービスの開発を進めている。どのような LOD をどのような形式でユーザに提示するかというポリシーを設定ファイルとして記述できるようにすることで、ユーザは興味のあるテーマに関する情報を、さまざまな LOD から容易に抽出し、それぞれの情報の特徴に合った形で閲覧することが可能となる。

3.1 設計思想

Open Knowledge Dictionary は、インターネットブラウザを利用した Web 閲覧中に、ユーザが選んだキーワードに関連する情報を簡単な操作で得られることを目的としたサービスである。このことから本サービスは、独立したアプリケーションではなく、Google の Chrome ブラウザのアドオンとして実装することとした。ユーザはまず、Open Knowledge Dictionary アドオンがインストールされた Chrome 上で、興味のあるキーワードをマウスで選択する。次にコンテキストメニューから



図 2: マイ健康ディクショナリサービス

Open Knowledge Dictionary を選ぶと、選択したキーワードのそばに、個々の LOD データを表す LOD 名ボタンが並び、ユーザがその中から 1 つを選択すると、該当する LOD からキーワードに合致する情報を検索し、その結果を LOD 名ボタンの脇にリスト表示する。

LOD からの情報取得方法は LOD サイトによってさまざまであり、情報抽出の際にはその違いを考慮した処理が必要である。本サービスでは、個々の LOD に合わせた取得方法および提示方法を、サービスの設定ファイルとして定義できるようにすることで、JavaScript などのソースコード修正を行うことなく各種 LOD からの情報検索処理を実装できるようにしている。処理の流れを、図 3 に示す。

manifest.json はサービスが最初に参照する設定ファイルであり、処理に用いる他の設定ファイル、menuconfiglist.json、および menuxxxxx.json の場所が記述されている。menuconfiglist.json ファイルには、menuxxxxx.json (“xxxxx” の部分は任意に記述可能) ファイルのリストが記述されている。menuxxxxx.json ファイルは、利用する LOD ごとに存在する設定ファイルであり、それぞれの LOD の取得方法と表示方法を記述する。以降の節で、LOD の取得方法、取得したデータの提示処理について解説する。

3.2 LOD の取得方法

データを取得するための LOD へのアクセス方法は、LOD の提供形態によって大きく次の 3 つに分類される。

- SPARQL によるアクセス
SPARQL エンドポイントに対して SPARQL 文でのリクエストを行うことにより、目的のデータを取得する
- API によるアクセス
指定された WebAPI に、所定のパラメーターやキーワードを指定したリクエストを行うことにより、目的のデータを取得する

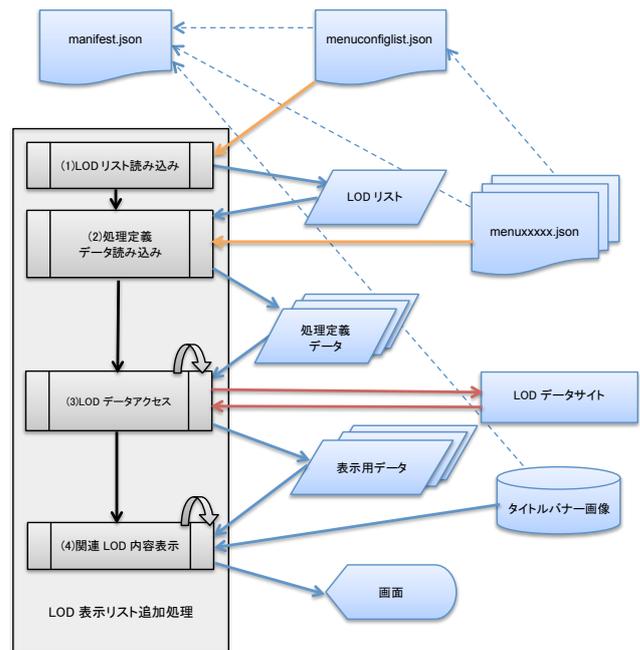


図 3: Open Knowledge Dictionary の処理の流れ

- データ取得によるアクセス
指定された WebAPI でファイルデータを取得し、そのファイルの中から目的のデータを抽出する

1 つめのタイプは、データサイト自体が SPARQL エンドポイントを整備している場合（例：日本語版 DBpedia, ヨコハマ・アート・LOD^{*3}）のアクセス方法であり、データを取得する際には、WebAPI の HTTP リクエストの中に SPARQL 文を内包し、目的の情報を JSON 形式などで得る。

2 つめは、データサイトに SPARQL 文を受け付ける機能が無い場合であり、データサイトが独自に規定したパラメーターなどを HTTP リクエストに設定して送る手法である。（例：ぐるなびレストラン検索 API^{*4}）。こちら、レスポンスを JSON 形式などで取得することができる。

3 つめは、RDF などのデータファイルそのものをデータサイトから取得し、そのファイルの中から、目的の情報を取り出す手法である。これは、データの公開などを目的とした「データカタログサイト」（例：LinkData.org^{*5} で公開されているデータ作品など）を対象とした場合に利用する手法である。

3.3 データ取得・表示処理

本節では、「SPARQL によるアクセス」によりデータを取得するタイプの LOD を対象とした処理の流れを例示する。Open Knowledge Dictionary サービスではまず、menuconfiglist.json を読み込み、検索対象となる LOD のリストを得る（図 3 (1)）。次に、LOD リストに基づき、各 LOD ごとの処理が記述された menuxxxxx.json ファイルを読み込み、処理定義データを得る（図 3 (2)）。

図 4 に処理定義データ menu.yart.json の例を、図 5 に Web ブラウザ上での LOD 情報の表示例を示す。この例は、ヨコハマ・アート・LOD からある特定のキーワードに関するイベントの情報を抽出し、表示する処理を表している。図 4 中の“endpointurl”はアクセスするエンドポイント、または WebAPI のアドレスを示している。“menuname”は表示の際の LOD 名ボタンのラベルである。“sparql”には実際に LOD サイトに送信する SPARQL 文の書式を記載する。

menuxxxxx.json 中に書かれている“RSV_”で始まるワードは予約語を表している。予約後の一覧を表 1 に示す。データ取得用の予約語は、ユーザが選択したキーワードや日付などを、データサイトからデータを抽出する際の条件として設定するために利用される。それらは menuxxxxx.json の、“sparqlplaceholder”の項目で指定され、SPARQL の文中では WHERE 句の中で指定される。例えば、menuxxxxx.json の“sparql”の項目で記述された SPARQL 文の中の RSV_KEYWORD の部分は、ユーザが選択したキーワードに置換される。その SPARQL 文がデータサイトへ送信され、得られたレスポンスから表示用のデータを生成する（図 3 (3)）。

情報表示用の予約語は、SPARQL リクエストで得られる結果の中から、実際に何をどのように表示するかを設定するために利用される。SPARQL で取得される“label”、“description”といった値と予約語の関係性は、menuxxxxx.json の“apimapkey”の項目において、マッピング情報として記述される。例えば、SPARQL 適用の結果得られるレスポンスの“label”の値は RSV_TITLE の中に入ることになる。

“headerform”、“titleform”、“titleplaceholder”、“descriptionform”、“descriptionplaceholder”、“additionalinfo-

form”、そして“addinfoplaceholder”は、SPARQL によって得られた情報のうち、実際に何をどの順番で表示するかを設定する部分である。例えば、“additionalinfoform”の中の“RSV_STARTPERIOD”、“RSV_ENDPERIOD”、“RSV_LOCATION”という部分は、それぞれ SPARQL のレスポンス中の、“dtstart”、“dtend”、“location”という変数に入っている値（文字列）に置き換えられ、その結果、「開催期間情報」の下に「場所情報」を出す、といったレイアウトで表示される（図 3 (4)）。このように、どの情報をキーに LOD データサイトから情報を抽出し、得られた結果をどのように表示するかを、1 つの JSON ファイルの中で記述することができる。

LOD の提供形態が「API によるアクセス」「データ取得によるアクセス」の場合に関しても、同様に menuxxxxx.json ファイルの中でその取得方法、および表示方法を指定することで、LOD の情報をブラウザ上で表示することが可能である。この設定ファイルに関しては、Open Knowledge Dictionary のサービスを実施する（アドオンを提供する）事業者が、ユーザが利用するであろう LOD データを想定して事前に提供してもよい。またユーザ自身がその設定ファイルを書き換えて、自分にとって有用な LOD データサイトおよびそこから得た情報の表示方法などをカスタマイズしてもよい。さらに、LOD データの提供者側で、Open Knowledge Dictionary 用の設定ファイルを公開する、という形態も考えられる。サービスの JavaScript などを修正することなく、検索対象の LOD サイトやその情報の表示方法を変えられるところが、Open Knowledge Dictionary の大きな特徴である。

3.4 インターフェース

Open Knowledge Dictionary の情報表示画面例を図 6 に示す。この例では、Web サイト上の「生涯学習」というキーワードを選択し、その関連情報を LOD から取得し、表示している。日本語版 DBpedia からはキーワードに関連する解説の情報を、国会図書館の LOD^{*6} からは関連する書籍データを、ヨコハマ・アート・LOD からは関連する芸術系イベントの情報を取得し、ユーザがそれらの中から所望の情報を選択して表示することが可能となっている。

また、Open Knowledge Dictionary のアドオンをインストールすると、Chrome のアドレスバーの脇に、Open Knowledge Dictionary ボタンが表示される。ユーザがキーワードをマウス選択していない状態でこのボタンを押下すると、本サービスは閲覧中の Web サイトからすべてのテキスト情報を抽出して形態素解析し、最も頻度の高い単語（名詞に限定）を検索キーワードとして LOD の検索処理を行う。このように、ユーザが検索キーワードを指定することなく、Web サイトのテーマに関連する情報を各種 LOD から検索し、提示することも可能となっている。

4. おわりに

本稿では、インターネットブラウザ上で Web サイトを閲覧している際に、ユーザが興味を持ったキーワードに関する情報を各種オープンデータから検索し、統合的に表示できるサービスを提案した。表示項目を設定するための予約語は、現在は想定するオープンデータの特徴に合わせて作成しているが、今後は schema.org^{*7} などの共通語彙の導入を検討し、選択されたキーワードが属するドメインに合わせた形で検索結果を表示す

*3 <http://yan.yafjp.org/lod/>

*4 <http://api.gnavi.co.jp/api/manual/restsearch/>

*5 <http://linkdata.org/>

*6 <http://www.ndl.go.jp/jp/aboutus/standards/lod.html>

*7 <http://schema.org/>

```

{
  "endpointurl": "http://data.yafjp.org/sparql?
  default-graph-uri=&format=PLH_FORMAT&timeout=0&query=",
  "encodetarget": [{"
    "PLH_FORMAT": "application/sparql-results+json"
  }],
  "menuname": "横浜アート情報",
  "menuid": "yartInfo",
  "responseType": "JSON",
  "method": "GET",
  "sparql": "
PREFIX rdf: <http://www.w3.org/.../22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX yav: <http://yafjp.org/terms/yav/1.0#>
PREFIX cal: <http://www.w3.org/2002/12/cal/icaltzd#>
PREFIX schema: <http://schema.org/>
SELECT DISTINCT
?label ?description ?dtstart ?dtend ?url ?location
WHERE {
  ?s a yav:Event;
  rdfs:label ?label;
  schema:description ?description;
  cal:dtstart ?dtstart;
  cal:dtend ?dtend.
  OPTIONAL { ?s schema:url ?url. }
  OPTIONAL { ?s cal:location ?location. }
  FILTER (xsd:dateTime(?dtstart) <=
  \\"RSV_TODAYISO8601\\" ^\"xsd:dateTime &&
  xsd:dateTime(?dtend) >=
  \\"RSV_TODAYISO8601\\" ^\"xsd:dateTime ).
  FILTER (contains(ucase(str(?description)),
  \\"RSV_KEYWORD\\") ||
  contains(ucase(str(?label)), \\"RSV_KEYWORD\\")).
  FILTER (lang(?label) =\"ja\" )
} ORDER BY ASC(?dtend) LIMIT 100",
  "sparqlplaceholder": [
    "RSV_KEYWORD",
    "RSV_TODAYISO8601"
  ],
  "apimapkey": {
    "RSV_TITLE": "label",
    "RSV_LINKURL": "url",
    "RSV_DESCRIPTION": "description",
    "RSV_STARTPERIOD": "dtstart",
    "RSV_ENDPERIOD": "dtend",
    "RSV_LOCATION": "location",
    "RSV_ADDITIONAL": ""
  },
  "bannerurl": "image/yokohama-art-lod-logo.png",
  "headerform": "<a href='http://yan.yafjp.org/lod'
  style='text-decoration: none; target='blank'>
  <img src='RSV_IMAGEPATH' width='209px'
  height='30px' class='sourceLogo'
  alt='ヨコハマ・アート・LOD'
  title='ヨコハマ・アート・LOD' /></a>",
  "nolinkurl": "http://yan.yafjp.org/lod",
  "titleform": "<a href='RSV_LINKURL' target='blank'
  style='text-decoration: none; >
  <span class='linkinmenu'>RSV_TITLE</span></a>",
  "titleplaceholder": [
    "RSV_TITLE",
    "RSV_LINKURL"
  ],
  "descriptionform": "RSV_DESCRIPTION",
  "descriptionplaceholder": ["RSV_DESCRIPTION"],
  "additionalinfoform": "RSV_STARTPERIOD
  ~RSV_ENDPERIOD<br>RSV_LOCATION",
  "addinfoplaceholder": [
    "RSV_STARTPERIOD",
    "RSV_ENDPERIOD",
    "RSV_LOCATION"
  ]
},
}

```

図 4: 処理定義データファイル (menu_yart.json) の記述例

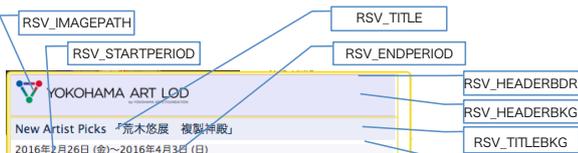


図 5: LOD データの表示例

る手法を検討する。さらに、閲覧対象の Web コンテンツの内容情報を深く解析することにより、コンテンツの理解をより促進する情報を外部のデータリソース群から検索し、それらを組み合わせ提示する手法の実現を目指す。今後は、提案システ

表 1: 主な予約語一覧

予約語名	意味
データ取得用	
RSV_KEYWORD	キーワード (指定単語) SPARQL 文で指定するためのキーワード
RSV_TODAYDATE	今日の日付 SPARQL 文で指定するための日付
RSV_TODAYISO8601	今日の日付 (ISO8601 形式) SPARQL 文で指定するための日付
情報表示用	
RSV_TITLE	タイトル
RSV_LINKURL	情報へのリンク URL
RSV_DESCRIPTION	内容
RSV_STARTPERIOD	開始日時
RSV_ENDPERIOD	終了日時
RSV_LOCATION	開催場所
RSV_ADDITIONAL	付加情報

図 6: Open Knowledge Dictionary の情報表示例 (出典: 文部科学省ホームページ)

ムの評価を行うとともに、対象とするコンテンツを Web サイトだけではなく、例えばテレビ番組のようなマルチメディアコンテンツにまで拡張し、多様なコンテンツの理解をサポートする手法の検討も行う予定である。

参考文献

[安蒜 10] 安蒜 孝政, 市村 光広, 佐藤 翔, 寺井 仁, 松村 敦, 宇陀 則彦, 逸村 裕: 図書館における情報探索行動, 2010 年日本図書館情報学会春季研究集会発表要綱 (2010)

[総務省 15] 総務省: 平成 27 年版情報通信白書 (2015)

[miyazaki 15] Miyazaki M., Urakawa M., Yamada I., Miura K., Miyazaki T., Fujisawa H., Nakagawa T.: My Health Dictionary -Study on Web Service using Program Information Data-hub as Linked Open Data-, Poster & Demo Track, The 14th International International Semantic Web Conference (2015)

[naseer 14] Naseer A., Kume T., Izu T., Igata N.: LOD for All: Unlocking Infinite Opportunities, The Semantic Web Challenge 2014, The 13th International International Semantic Web Conference (2014)