

非構造化データからの Linked Data 化および FIWARE に 対応した地域防災データ統合配信 API の試作

吉賀 夏子^{1,a)} 堀 良彰^{1,b)} 牛島 清豪^{2,c)}

概要 :

近年, 我が国を含む世界中で極端な気象・気候現象が発生している. 被災地では, 自治体等の避難勧告に加え, 個人が緊急事態発生時に即時的かつ出所の多様なリソースを取得し, 総合的に判断して自らの行動を決定する必要がある. 本稿では, 防災関連メール等の非構造化データを Linked Data に変換し, 標準化された情報を発行可能な FIWARE の枠組みを用いて, 地域防災データを統合配信可能なシステムの試作を行った.

Transforming Unstructured Data into Linked Data and Constructing a Prototype of FIWARE-ready API for Hyper-local Disaster Information

NATSUKO YOSHIGA^{1,a)} YOSHIAKI HORI^{1,b)} KIYOHIDE USHIJIMA^{2,c)}

Abstract: In recent years, extreme climate phenomena have been observed all over the world. People in disaster areas need to make a total decision about their own actions by getting real-time and diverse resources when an emergency happens. This paper describes converting unstructured data extracted from disaster prevention emails into Linked Data. The Linked Data are registered on FIWARE-enable system to unify diverse information such as hyper-local water level data which are also converted into Linked Data.

1. はじめに

2018 年・2019 年は, 世界各地で極端な気象・気候現象が発生した. 我が国においても, 西日本から東海地方を中心に広い範囲で数日間大雨が続き (平成 30 年 7 月豪雨), 全国の降水量の総和は, 1982 年以降の豪雨災害時の降水量の中で最も多い値となった [1].

平成 30 年 7 月豪雨の際, 多くの被災地では自治体から避難勧告が発令される等, 避難行動を促す情報が出されたものの, 自宅等に留まり適切な行動を取ることができず, 多くの方が亡くなる結果となっている. このような事態を踏まえ, 中央防災会議防災対策実行会議の下に設置された

「平成 30 年 7 月豪雨による水害・土砂災害からの避難に関するワーキンググループ」では, これまでとは異なり, 住民が「自らの命は自らが守る」意識をもって自らの判断で避難行動をとり, 行政はそれを全力で支援するという, 住民全体の取組強化による防災意識の高い社会の構築に向けた報告が取りまとめられている [2]. その報告の中で, 市町村は空振りをおそれずに躊躇なく避難勧告等を発令することを基本とし, 発令の対象を明確にするとともに, 対象者がとるべき避難行動がわかるように繰返し伝達するとの方針が定められている [2].

防災のために, 住民は自ら適切な行動を起こすための情報を速やかに入手できることが理想である. しかし, 自治体や国が提供している情報は, 電子メールや人間が参照することを前提とした非構造化データ (メッセージ) であるものが多い. 筆者らは, これまで水防情報の速やかな伝達のために河川水位を含む水防情報の構造化を検討した [16].

本稿では, 水防情報の構造化のためのオントロジー作成

¹ 佐賀大学

Saga University, Saga shi, Saga 840-8502, Japan

² 株式会社ローカルメディアラボ

Local Media Labs, Saga shi, Saga 840-0804, Japan

a) natsukoy@cc.saga-u.ac.jp

b) horiyo@cc.saga-u.ac.jp

c) saygo.ushijima@lm-labs.com

と Linked Data 化の知見を基に、自治体をはじめとする公共機関等から提供される、電子メール形式の非構造化メッセージの自動構造化手法およびそれを適用して得られた構造化データの蓄積と参照のためのオントロジー生成と Linked Data 共有 API の構築について提案する。

2. 地域に散在する防災情報の収集、蓄積および配信

2.1 概要

地域限定の防災に関連する情報は、Web サイト上に多様な形式で散在しており、その多くは非構造化形式であるため、知識として蓄積・利用が困難な側面を持つ。また、自治体等が発信する多くの防災情報が、自然文による内容で占められている。そのため、情報を即時的に利活用するために必要な機械可読化処理が施されているデータが少ないことが課題である。

多様なデータを機械可読化する一連の技術あるいはデータセットを Linked Data と呼ぶ。Linked Data 化されたデータセットは、データ項目の定義まで Web で参照可能な構造を有するテキストデータであるため、他者と共有しやすく機械処理が容易であるためシステム構築の際にアプリケーション部とデータセット部を分離できる。そのため、データセットのみ長期保存したり、異なるアプリケーションに適用したりするといった、柔軟な対応が可能となる。

しかし、本稿執筆時点において、Web 上での語彙定義およびデータ項目の構造化を行うオントロジー [7] 技術を用いて、実際に日本語で記述された防災関連の Linked Data を構築し、データの配信までを可能にした事例を、筆者らは見つけることができなかった。

一方、Linked Data 化されたデータセットを、共有し様々なアプリケーションから参照するための枠組みの研究開発が進められている。ETSI (欧州電気通信標準化機構) では、Linked Data に対応したコンテキスト情報管理のための API である NGSI (Next Generation Service Interface) の標準化 NGSI-LD [6] が行われている。NGSI-LD では、多くの開発者によく知られている Linked Data のシリアル化表現形式 (JSON-LD) を使用してデータを送信あるいは要求する方法が定義されている。データおよびそのデータの意味、関係、ソース、ライセンス等のコンテキストを同時に取扱うことを可能にする。

商業用途のみでなく、公共利用も視野に入れたデータプラットフォーム実現のための FIWARE [3] プロジェクトがある。FIWARE プロジェクトでは、NGSI-LD を実装したデータプラットフォームとして、Linked Data [4] 対応のコンテンツブローカー実装 Orion-LD [11] が行われている。なお、FIWARE プロジェクトでは、これまで NGSI version 1 および同 version 2 を実装した Orion コンテンツブロー

カーを開発しており、日本においても兵庫県加古川市 [5] 等で公共情報の共有が開始されているが、これらは Linked Data に対応した NGSI-LD の前段階である。

これまで Linked Data に基づくデータを API 経由で公開する場合、データ提供者が API も含めて独自に構築する必要があった。この場合、一定以上の品質の API の設計は、技術面やコスト面で敷居が高いものとならざるを得なかった。そこで、筆者らは、Linked Data を取扱うことを可能にするデータプラットフォームの構築においては、標準化された枠組みである NGSI-LD 仕様に基づく実装 Orion-LD を導入することで、データ共有とその利用のための一定の品質を有する API 公開を比較的容易に実現できると考えた。

本稿では、佐賀県で防災・治安情報を逐次配信している「防災メールあんあん [8] (以下、メールと呼ぶ。) を対象に、メール内容を NGSI-LD に対応した Linked Data に変換して蓄積し、動作検証した。その際、著者が開発した Linked Data 自動変換システムの形態素解析および固有表現抽出プロセスを一部応用した。このシステムは、元々、江戸時代の業務日誌を蓄積する「小城藩日記データベース」[9] 中の記事文を自動で機械可読化するために構築したものである。

2.2 FIWARE および NGSI-LD の概要

データは再利用されてこそ価値がある。W3C (The World Wide Web Consortium) は、Web 空間に存在するデータセットに対し、再利用可能かの観点から 5 段階で格付けする指標を提案している [10]。

具体的には、1) データセット自体が Web から閲覧可能か、2) 誰でも利用可能な許諾か、3) すぐにアプリケーションに使用できる形式で用意されているかを指標としている。これらのうち、最後に挙げたソフトウェアに使用できる形式については、データセット中の項目について構造化された定義が、どの程度共有可能かが重要である。項目の関係定義が明確かつ機械可読であれば、アプリケーションにデータを正確に投入できる。

これまでに、FIWARE 財団 (FIWARE Foundation) [3] では、商業用途のみでなく公共利用も視野に入れたデータプラットフォームとして、FIWARE と呼ばれる一連のオープンソース開発プロジェクトを実施している。これらは、先に述べた ETSI での標準化活動の成果を取り入れている。そのうち、多様な形態をしたデータセットから、日時、場所、主題になりうる物事等の関係を抜き出して、同じプラットフォームに納めるコンテキスト情報管理 (Context Information Management) についての標準化が進められている。

2019 年 1 月には、データセット格付けで最上位にあたる

構造化データを、FIWARE 標準を通じて利用促進させるために、項目の意味情報までも Web で取り扱える Linked Data ベースのコンテキスト情報管理 API NGSI-LD (ETSI GS CIM 009) (2020 年 2 月における最新仕様は、version 1.2.2 [6]) が策定された。

NGSI-LD の情報モデルは、Web 上で外部組織で構築されたオントロジーを再利用することを前提に、全てのドメインに共通の Core Meta Model, 複数のドメインに共通のオントロジー, および各分野特有のドメインオントロジーの三層に分けて形成されている。これらのオントロジーは、NGSI-LD API に実データを投入する際に、コンテキストデータとして実データに付記される。

NGSI-LD に準拠した FIWARE Orion-LD のアーキテクチャは、コンテンツブローカーを中心とする集中型である。基本的に、コンテンツブローカーでは、データ登録や更新を行うコンテンツプロデューサーと、登録された情報をクエリ等で選択的かつ能動的に受け取ることが可能なコンテキストコンシューマの動作を管理する。

アプリケーションは、NGSI-LD API を用いることで、コンテンツブローカー機能を持つコンテンツサーバに対し、Linked Data を表現可能な構造化データ形式のひとつである JSON-LD を利用したデータ操作 (POST, GET, DELETE 等) を実現する。

次節以降に、具体例として、メールの内容を Linked Data 化し、NGSI-LD 仕様準拠した Orion-LD[11] に蓄積する手順を述べる。

2.3 メール内容の Linked Data 化

2.3.1 項目の定義および構造化

防災メールの内容を Linked Data 形式に構造化するため、まず、データ構造を定義するオントロジーを設計する。そのため、Web 空間で一意に指定可能な URI, 関係する項目同士の関係および後述の固有表現と防災メール間の関係を、図 1 に示すような ER 図を作成して定義する。

記事クラス 本クラスのオブジェクトは、防災メールの構成情報である。防災メールの取得元、件名、登録日、本文、id、type のデータプロパティを持つ。件名と本文については、総文字数および内容を示す値で構成される。また、このクラスは 0 個以上の固有表現で構成されることを示すため、オブジェクトプロパティ hasNamedEntity で固有表現 (クラス) を参照する。

固有表現 本クラスのオブジェクトは、固有表現の情報である。固有表現は、内容を示す値、件名あるいは本文中かを示す所在、文字位置 (開始および終了位置)、id、type のデータプロパティを持つ。また、固有表現は後述する“固有表現クラス名” (表 1) のクラスを refNamedEntityClass を参照する。加えて、固有表現

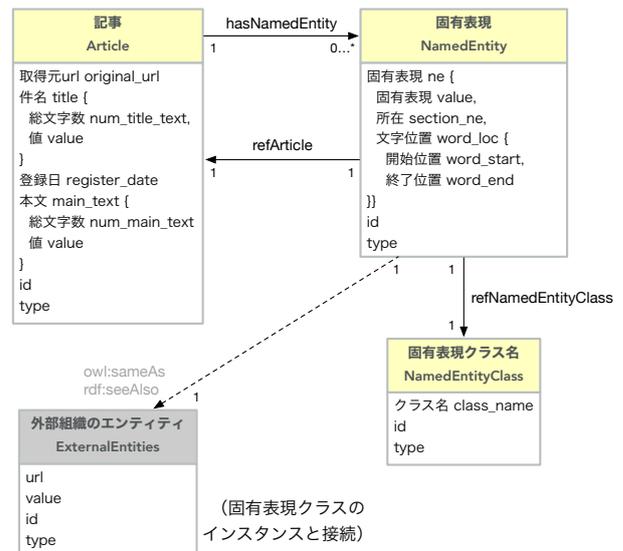


図 1 防災メールの ER 図

Fig. 1 An entity-relationship model of the hyper-local disaster prevention email.

クラスから記事クラスを refArticle で参照し、固有表現の出所を明確にする。

固有表現クラス名 本クラスのオブジェクトは、日時、地名、トピック、組織名、数、その他のクラス名で構成されている。

外部組織のエントティ 参考として、固有表現クラスのインスタンスには、外部組織のインスタンスを Web 上の様々な組織が定義しているオブジェクトプロパティ (メタデータスキーマ) で接続できることを示す。例えば、ある固有表現が owl:sameAs で Wikidata 項目と同義、rdf:seeAlso で別途参照可能であることを表現できる。

図 1 を基に、定義したデータ項目とその構造について、URI 参照解決できるように、意味の定義を OWL (Web Ontology Language)[7] で構造化し、Web アクセス可能なオントロジー定義ファイルを作成する*1。オントロジーのクラスおよびクラス間の関係 (オブジェクトプロパティ)、クラスを構成する項目 (データプロパティ) 定義には、OWL 作成ツール Protégé[12] を用いた。

定義したデータ項目については、外部組織のデータ項目との共通点および相違点を検討し、対象データ独自のデータ項目およびそのクラスを、分野限定的なドメインオントロジーと外部組織から引用可能な共用オントロジーに仕分ける。一般に、共通のデータ項目については、ETSI や Wikidata のような、可能な限り標準化が行われ、管理が長期間にわたって維持できているものを再利用する。

NGSI-LD では、ドメインオントロジーと共用オントロジー定義データを、実際の値が入った Linked Data に後述

*1 著者が作成した防災メールオントロジー: <https://github.com/nikolito/anand-ld>

表 1 固有表現クラスの内訳および各クラスの単語例.

Table 1 Named entity classes of the hyper-local disaster prevention email and the descriptions.

クラス名	クラスのラベル名	説明
日時	DATE	日時を表す名詞
地名	PLACE	地図上で特定可能な場所
トピック	TOPIC	文中の内容を表現し得る名詞
組織名	ORGANIZATION	個人名以外の名称
数	NUMBER	数詞
その他	ETC	上記以外の定型句

のコンテキストデータに紐づけることで、API の発行時に Web で参照解決可能なデータ項目として取り扱える (図 2).

2.3.2 構造化された定義に沿った防災メールデータの Linked Data 化

前節で定義したオントロジーに沿って、防災メールのデータを Linked Data 化する. Linked Data 化に際し、著者が開発した Linked Data 自動変換システム [15] を、防災メール用に対応させたものを利用した. 図 2 にその概要を示す.

まず、Linked Data 化システムの入力について述べる. Web で蓄積されているデータもしくは直接受信したメールから、スクレイピングを行い、送信日、件名、本文等の項目に整理した JSON データに変換する (図 3 上).

他方、件名および本文から、検索の際にキーワードになりうるフレーズを固有表現として抽出するために、形態素解析ツール MeCab[13] を用いる. 形態素解析では、自然文を意味のある単語に分割し、文法情報を得られる. MeCab は、形態素解析で広く使用されているため導入が比較的容易である上、利用者側で、文法情報中のデータ項目や地域固有の単語をユーザ辞書として登録可能である.

そこで、図 3 下のように、MeCab での自然文解析結果に固有表現クラスを組み込むことで、防災メール中の固有表現を日時、場所、関与した組織、人名等の大まかな意味付けを行うクラスに分別可能にする. 各固有表現クラスの内訳は表 1 に示す.

抽出候補の選定には、現代文から固有表現抽出可能なツールの COTOHA[14] を用いた. 加えて、佐賀県の地名、官公庁の組織名等の固有名詞を収集した. これらの固有名詞は、表 1 に示す固有表現クラスごとに仕分けたファイルにあらかじめリストする. これらのファイルは、MeCab のユーザ辞書作成に用いられる (図 2 左).

次に、自然文から抽出した固有表現を構造化について述べる. 例えば、メール内容であれば、送信日時、発信者、件名、本文等の構成要素にクラス分けし、さらにそのクラスを構成するデータ項目とその型を定義する. データ項目の定義については、Web で一意の URI (Unified Resource Identifier) を設定することで、意味説明および関係情報を

Web 上で機械的に参照可能にする. これにより、ある組織のデータセットと別の組織で定義された属性の関係について同義なのか、上位下位属性なのか等の関係性について、アプリケーションによる機械的な取り扱いが容易になるため、Linked Data の利点を生かすことができる.

最後に、スクレイピングした JSON 形式のメールデータおよび NGSI-LD 対応コンテキストデータを参照する URI を用いて、Linked Data を自動生成する (図 2 中央下). 本自動作成過程において、入力したメールデータを MeCab で防災メールの ER 図の通りに構造化された JSON-LD 形式の Linked Data に変換する*2 (図 2 右下).

2.4 Linked Data 化された地域情報の蓄積および配信

Linked Data 化された防災メールのデータは、生成と同時に、Linked Data ベースのコンテキスト情報管理 API である NGSI-LD に HTTP/HTTPS を通じてアップロードされる. NGSI-LD は、アップロードされたデータのコンテキスト情報を読み、データ項目 URI の有効性までを評価し結果を返す.

本稿では、NGSI-LD API を Linked Data に対応させるために、FIWARE Orion Context Broker を機能拡張した Orion Context Broker (Linked Data Extensions)[11] (以下、Orion-LD) を使用した. Orion-LD の実装は、Docker を用いて行った.

NGSI 規格の API を採用することで、防災メール情報に加えて、別途開発した水防情報のオントロジー [16] *3、コンテキスト情報、実際の水防データ等異なる種類の情報も同じ API の仕組みでデータ蓄積と配信が可能となる.

3. おわりに

本稿では、自治体等が発信する多くの防災情報が、自然文による内容で占められているため、実質的に、情報を Web で機械的かつ即時的に利活用できないことに着目した.

具体的には、多様なデータ配信を行う標準の枠組みである FIWARE のもとで策定された、NGSI-LD API 互換の Orion-LD を用いて、現在メールで配信中の防災関連情報を Linked Data として蓄積し、アプリケーションに配信するシステムについて検討した. その際に、メール内容をもとに、防災メールオントロジーを設計し、実データをオントロジーに沿って変換した. その後、NGSI-LD にメールデータを蓄積し配信できることを確認した.

その結果、佐賀県の河川状況を記録している水防データについても同様に、NGSI-LD API を蓄積・参照可能であることから、多様で異なる組織から配信されたデータを Linked Data に変換することで、同じ標準 API を用いて運

*2 <https://github.com/nikolito/anand-lid/> に Linked Data 生成例を記載.

*3 https://github.com/nikolito/suibou-lid

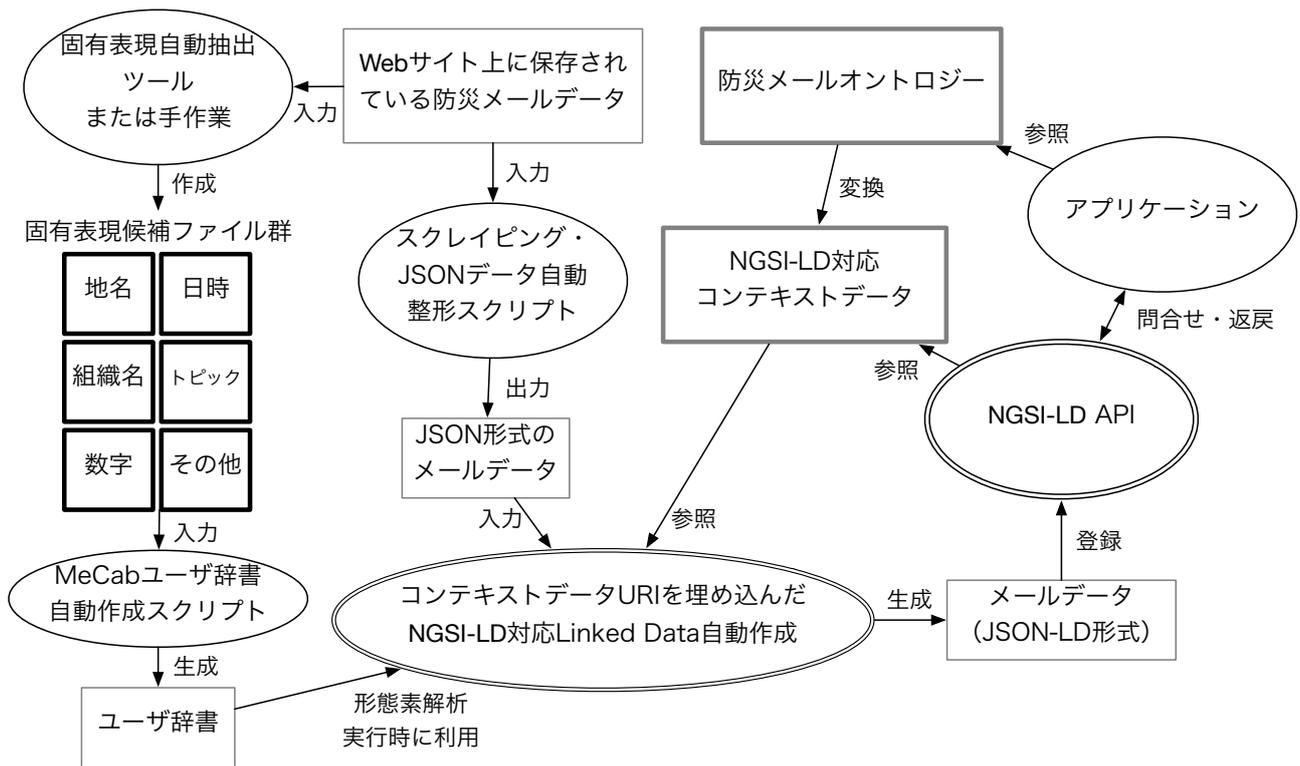
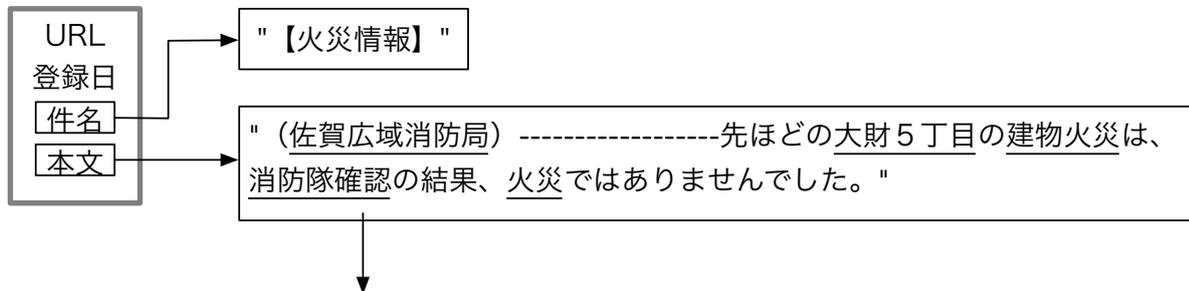


図 2 防災メールデータを入力として Linked Data を出力し，Orion-LD API に蓄積および配信するシステムの模式図

Fig. 2 A model of the system converting Linked Data from disaster prevention emails, accumulating and delivering those data.

防災メール



MeCabによる形態素解析の結果（固有表現のみ抜粋）

佐賀広域消防局,名詞,固有名詞,組織,*,*,*,*,佐賀広域消防局,*,佐賀広域消防局,*,固,*,*,*,ANAN_ORG
 大財5丁目,名詞,固有名詞,地域,一般,*,*,*,大財5丁目,,大財5丁目,,固,*,*,*,ANAN_PLACE
 建物火災,名詞,固有名詞,一般,*,*,*,建物火災,,建物火災,,和,*,*,*,ANAN_TOPIC
 消防隊,名詞,固有名詞,一般,*,*,*,消防隊,,消防隊,,和,*,*,*,ANAN_TOPIC
 確認,名詞,固有名詞,一般,*,*,*,確認,,確認,,和,*,*,*,ANAN_TOPIC
 火災,名詞,固有名詞,一般,*,*,*,火災,,火災,,和,*,*,*,ANAN_TOPIC

図 3 Linked Data 化システムでの防災メール入力例。形態素解析結果で「ANAN_***」とラベルが付与された単語が固有表現。

Fig. 3 An example of a disaster prevention email on the system converting to Linked Data. Each word which labeled as “ANAN_***” represents as a named entity.

用可能であることを示した。

今後は，防災のための異種情報参照アプリケーションの

設計と実装を進めたい。

謝辞 本研究の一部は，2019年度総務省戦略的情報通信

研究開発推進事業（SCOPE）委託研究の成果である。本研究の一部は、日本学術振興会科学研究費補助金（JSPS KAKENHI Grant Number JP19K20630）の成果である。

参考文献

- [1] “気候変動監視レポート 2018”, 気象庁, July 2019
- [2] “避難勧告等に関するガイドライン ①（避難行動・情報伝達編）”, 内閣府（防災担当）, 平成 31 年 3 月
- [3] FIWARE, The FIWARE Foundation. <https://www.fiware.org> (lastaccess:2020-02-21)
- [4] Tom Heath and Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space (1st edition)*, Morgan & Claypool, 2011
- [5] オープンデータ（API）を活用する開発者の方へ / 加古川市 <https://www.city.kakogawa.lg.jp/soshikikarasagasu/kikakubu/jouhouseisakuka/opendata/1528499081092.html> (lastaccess:2020-02-23)
- [6] NGS-LD API, INDUSTRY SPECIFICATION GROUP (ISG) CROSS CUTTING CONTEXT INFORMATION MANAGEMENT (CIM) 009, version 1.2.2, ETSI, <https://www.etsi.org/committee/cim> (lastaccess:2020-02-21)
- [7] OWL 2 Web Ontology Language Primer (Second Edition), W3C Recommendation, December 2012 <https://www.w3.org/TR/owl-primer/> (lastaccess:2020-02-21)
- [8] 「防災ネットあんあん」で配信している内容について紹介します <http://www.pref.saga.lg.jp/kiji0031196/index.html> (lastaccess:2020-02-23)
- [9] 小城藩日記データベース <https://www.dl.saga-u.ac.jp/ogiNikki/> (lastaccess:2020-02-23)
- [10] 5 Star Linked Open Data <https://www.w3.org/TR/ld-glossary/#x5-star-linked-open-data> (lastaccess:2020-02-23)
- [11] FIWARE Orion Context Broker (Linked Data Extensions) <https://github.com/FIWARE-GEs/orion-ld> (lastaccess:2020-02-21)
- [12] Protégé, A free, open-source ontology editor and framework for building intelligent systems <https://protege.stanford.edu> (lastaccess:2020-02-22)
- [13] MeCab: Yet Another Part-of-Speech and Morphological Analyzer <https://taku910.github.io/mecab/> (lastaccess:2020-02-23)
- [14] COTOHA API. <https://api.ce-cotoha.com/contents/index.html> (lastaccess:2020-02-23)
- [15] 吉賀夏子, 只木進一. “古典籍書誌データ構造に対応した linked data への半自動変換”. 情報処理学会論文誌, 59(2), 257-266, 2018
- [16] 吉賀夏子, 堀良彰, 牛島清豪. “防災のための Linked Data を用いた河川水位情報の蓄積配信システムに関する検討”, 2019 年度（第 72 回）電気・情報関係学会九州支部連合大会, 2019