

新潟オープンデータハブ企画書

本サービスで実現したいこと

つながりことでデータの価値を高める。

新潟に関連するオープンデータをつなげて言葉で簡単に抜き出せるようにする。

AsIs

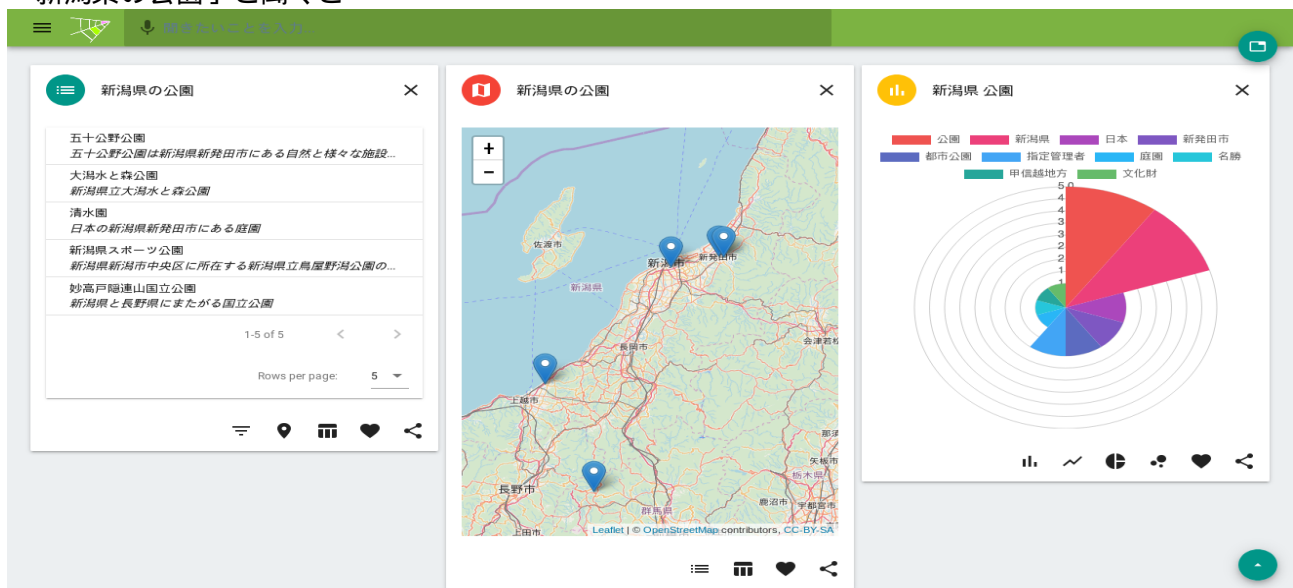
現状、オープンデータの多くは、CSV やエクセルファイルの形で提供され、カテゴリ別もしくは、全文検索でたどり着くことが出来ます。

ToBe

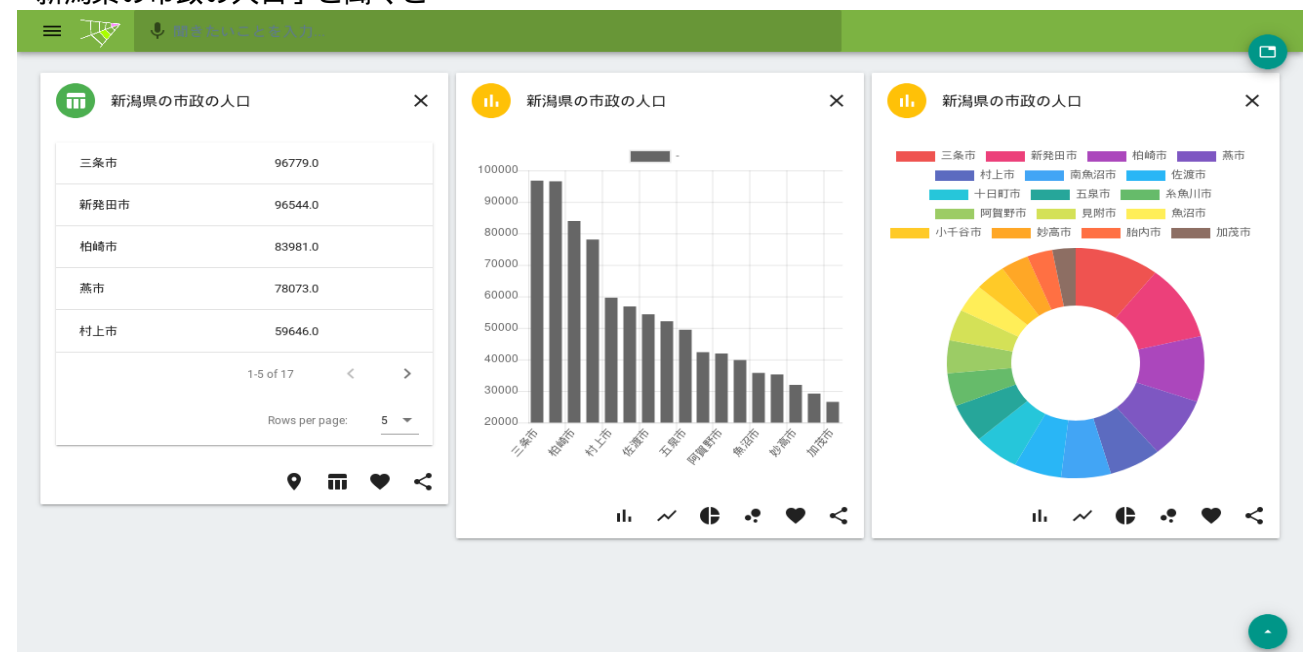
新潟に関するオープンデータを一箇所で蜜につなぐ。欲しい情報を直感的に言葉(音声入力を含む)で抜き出せるようにして、より多くの人々が有効活用できるようにする。

オープンデータ活用イメージ

「新潟県の公園」と聞くと



「新潟県の市政の人口」と聞くと



「新潟県の政治家の出生地」と聞いて、「湯沢町」→「出生地」とリンクを辿ると湯沢町出身の別の人もわかります

The screenshot displays a grid of nine data panels from a web application. The panels are arranged in three rows and three columns. The top row shows a list of politicians from Niigata Prefecture, a bar chart of birthplaces, and a profile for Yuzawa Town. The middle row shows a map of Yuzawa Town, a circular diagram of administrative divisions, and a classification tree for Yuzawa Town. The bottom row shows a list of politicians born in Yuzawa Town, a profile for Takano Yuzo, and a profile for Takano Tomoko.

新潟県の政治家の出生地

栗原博久	新潟県
上村清隆	湯沢町
小池清彦	加茂市
会田洋	新潟県
田村正幸	湯沢町

1-5 of 89
Rows per page: 5

新潟県の政治家の出生地

Bar chart showing birthplaces: 新潟県 (45), 柏崎市 (10), 西區 (10), 神林村(新潟県) (10), 小千谷市 (10).

湯沢町
日本の新潟県の町

名前: 越後湯沢
分類: 日本の町
分類: 市区町村
公式ウェブサイト: <https://www.town.yuzawa.lg.jp/>
国: 日本
1-5 of 11
Rows per page: 5

湯沢町

Map showing Yuzawa Town location in Niigata Prefecture.

湯沢町

Legend: 位置する行政区画 (red), 直下の行政区画 (purple), 起終点 (pink), 出生地 (yellow).
Circular diagram showing administrative divisions.

湯沢町

Classification tree: 湯沢町 (yellow star) → 分類 → 日本の町 (blue square).

湯沢町 出生地

岸野裕
たかの友梨
田村正幸
上村清隆

1-4 of 4
Rows per page: 5

岸野裕
日本の実業家

名前: 岸野裕
分類: ヒト
性別: 男性
出生地: 湯沢町
生年月日: +1964-04-01T...

1-5 of 10
Rows per page: 5

たかの友梨
エステティシャン、アンチエイジン...

名前: 高野友梨
分類: ヒト
性別: 女性
出生地: 湯沢町
生年月日: +1948-01-22T...

1-5 of 10
Rows per page: 5

他にも出来るようになること

- 「新潟市のIHヒーターのごみの分別区分」と聞くと「燃やさないごみ」とダイレクトに回答
- 「新潟県のすいかの残留農薬の検査結果」と聞くと「基準値超過なし」と回答
- 「新潟県で残留農薬の規格基準違反」と聞くと該当する食品の一覧を表示
などなどインポートするデータによって出来ることが増える

⇒一箇所にデータが蜜につながっているので、聞いたことに対して直接回答することも出来るし、一覧を回答することも出来る。特定の分野の専用アプリケーションでないので、新たなデータが増えてもインポートするだけで、柔軟に対応できる。観光、暮らし、福祉、交通、行政など様々な場面でオープンデータの価値を高めて、利活用することが出来ます。また、本システムで回答が導けなかった質問に対しては、全文検索エンジンにフォールバックし、その回答を返すことで今提供されているサービスレベルは維持することが出来る選択肢も準備出来ます。

具体的な方策

データ

以下のようなデータが手元にあったとする。

名前	時期	科目	点数
A 君	夏学期期末	数学	80
A 君	夏学期期末	英語	60
B 君	夏学期期末	数学	100
B 君	夏学期期末	英語	80

名前	性別	生年月日	出身地
A 君	男	2019年5月	〇〇市
B 君	女	2019年6月	××市

アップロード

エンティティ化する際の key をユーザが指定してアップロード。もしくは、一意な key になるように自動で判断する。

※前者なら、「名前-時期-科目」を key にする。後者なら「名前」を key にする。

構築された RDF

列を一つの主題とした、以下のような RDF が構築される。

```
<A 君-夏学期期末-数学> <名前> <A 君> ;
    <時期> <夏学期期末> ;
    <科目> <数学> ;
    <点数> <80> .
<A 君-夏学期期末-英語> <名前> <A 君> ;
    <時期> <夏学期期末> ;
    <科目> <英語> ;
    <点数> <60> .
...
...
...
<A 君> <名前> <A 君> ;
    <性別> <男> ;
    <生年月日> <2019年5月> ;
```

```
<出身地> <〇〇市> .
<B 君> <名前> <B 君> ;
  <性別> <女> ;
  <生年月日> <2019 年6 月> ;
  <出身地> <××市> .
```

自然言語クエリ

自然言語で簡単にデータを操作できるようになる。

「A 君の数学の点数」→A 君の数学に関する点数が一覧表示される

「夏学期期末の数学の点数」→全員の夏学期期末の数学に関する点数が一覧表示される

「〇〇市の数学の点数」→〇〇市出身の全員の数学の点数が一覧表示される

「△△県の数学の点数」→△△県出身の全員の数学の点数が一覧表示される

※操作イメージは wikidata のデータを利用した実装版である <https://app.wiev.net> を参照。

SPARQL クエリ

更に詳細にデータを操作したい場合は、SPARQL で操作することも可能。

つながることの利点

- データ量が増えるに従って増えるファイル、ページを探す手間を省きダイレクトに欲しいデータにアクセスできる
- 通常のグラフはもちろん、マップ、力学グラフなどグラフィカルな表示を自動で付加し直感的にデータを扱える

課題設定

データを集める

- 県や市など各自治体で提供しているデータとそのメタデータを集める
 - 各ドメインに合わせたクローラーでリンク名とセットでファイルを収集する(CSV、XLS ファイルを想定)
 - 可能であれば付加的なメタデータも収集する
- 集めたデータを自動で RDF 化する
 - リンク名、ファイル名、付加メタデータ、ファイル内容から、一意となるような key を合成し、各行毎にエンティティ化し、各列、及びファイル名等と関係を構築する
 - ファイル名は自然言語処理で単語毎に分割し、関係の素とする

データをつなげる

- RDF を一箇所の RDF ストアでつなげる

簡易なデータ抽出

- 自然言語でつながりを自由に簡単に抽出する