

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

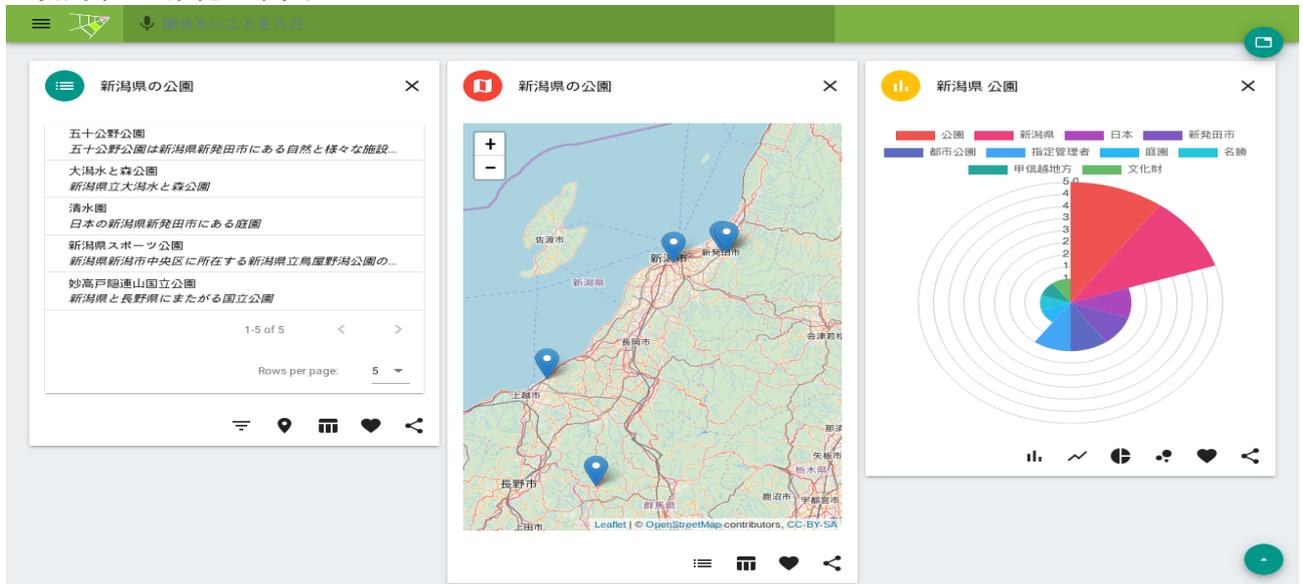
現状、オープンデータの多くは、CSV やエクセルファイルの形で提供され、カテゴリ別もしくは、全文検索でたどり着くことが出来ます。

ToBe

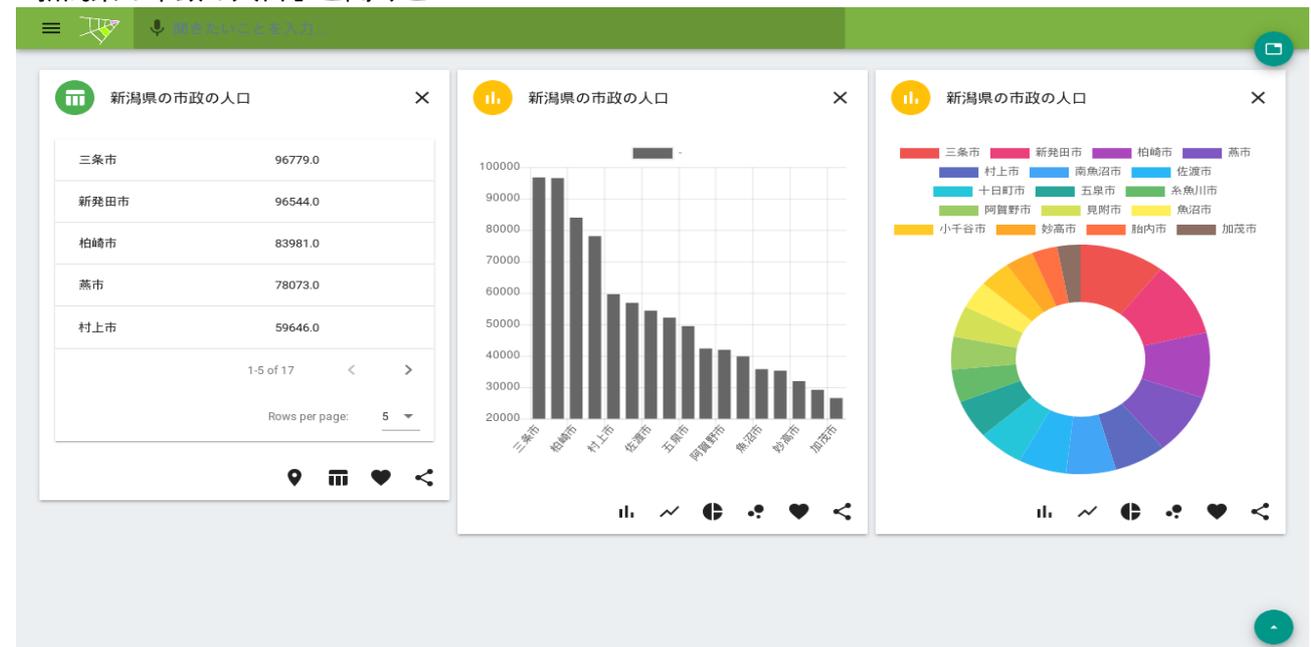
新潟に関するオープンデータを一箇所で蜜につなぐ。欲しい情報を直感的に言葉(音声入力を含む)で抜き出せるようにして、より多くの人々が有効活用できるようにする。

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



「北陸道のサービスエリア」と聞くと

☰
🗺️
🔊
🗎

☰
北陸道のパーキングエリア
✕

- 宋パーキングエリア
日本の新潟県三条市にある北陸自動車道のパーキングエ...
- 越中垣パーキングエリア
日本の富山県下新川郡朝日町にある北陸自動車道のパー...
- 高岡パーキングエリア
日本の富山県高岡市にある北陸自動車道のパーキングエ...
- 大潟パーキングエリア
日本の新潟県上越市にある北陸自動車道のパーキングエ...
- 奥羽パーキングエリア
日本の富山県射水市にある北陸自動車道のパーキングエ...

1-5 of 19

Rows per page: 5

☰
📍
🗺️
📄
❤️
🔗

📍
北陸道のパーキングエリア
✕



☰
📄
❤️
🔗

「新潟県政治家の出身校」と聞くと

☰
🗺️
🔊
🗎

☰
新潟県政治家の出身校
✕

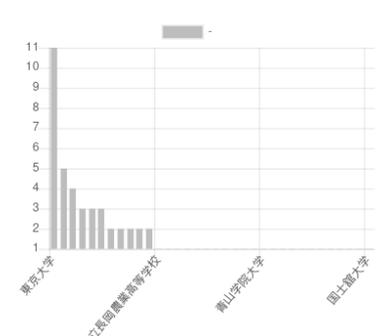
小林清	新潟県立巻高等学校
斎藤洋明	神戸大学
塚田十一郎	一橋大学
花角英世	東京大学
斎藤洋明	学習院大学

31-35 of 62

Rows per page: 5

📍
🗺️
📄
❤️
🔗

📊
新潟県政治家の出身校
✕



出身校	人数
東京大学	11
新潟県立巻高等学校	5
新潟学院大学	4
国士舘大学	3
神戸大学	2
一橋大学	2
学習院大学	2
京都府立総合技術学院	2
富山大学	1

📊
📄
❤️
🔗

☰
新潟県政治家の出身校 新潟大学
✕

栗原博久
荒井賢太郎
西村智奈美
森裕子

1-4 of 4

Rows per page: 5

☰
📍
🗺️
📄
❤️
🔗

「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 organized as follows:

- Top Row:**
 - Panel 1:** 「新潟県の政治家の出生地」 (Birthplaces of politicians in Niigata Prefecture). A table lists names and birthplaces, with 「湯沢町」 (Yuzawa Town) highlighted in red.
 - Panel 2:** 「新潟県の政治家の出生地」 (Birthplaces of politicians in Niigata Prefecture). A bar chart showing the distribution of birthplaces across various municipalities.
 - Panel 3:** 「湯沢町」 (Yuzawa Town). A profile card for the town, including its name, classification, and official website.
- Middle Row:**
 - Panel 4:** 「湯沢町」 (Yuzawa Town). A map view of the town's location.
 - Panel 5:** 「湯沢町」 (Yuzawa Town). A circular diagram showing the town's position relative to other administrative areas.
 - Panel 6:** 「湯沢町」 (Yuzawa Town). A vertical flow diagram showing the classification hierarchy from 「湯沢町」 to 「日本の町」 (Municipalities in Japan).
- Bottom Row:**
 - Panel 7:** 「湯沢町 出生地」 (Birthplaces in Yuzawa Town). A table listing names and birthplaces, with 「岸野裕」 (Kishino Hiro) and 「たかの友梨」 (Takano Tomori) highlighted in red.
 - Panel 8:** 「岸野裕」 (Kishino Hiro). A profile card for the politician, including his name, classification, gender, birthplace, and birth date.
 - Panel 9:** 「たかの友梨」 (Takano Tomori). A profile card for the politician, including her name, classification, gender, birthplace, and birth date.

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