

発想支援とジオプロセシングのシームレスな統合に向けた QGIS プラグインの開発

富山県立大学工学部電子・情報工学科
1715059 平松楓也

情報基盤工学講座
指導教員：奥原浩之

1 はじめに

昨今、世界規模で流行しているコロナウイルスの感染状況や人口、降水量、地層などの異なる情報を地図上に重ねることで、それまで見えなかったものを可視化するソフトとして、QGIS などの GIS ソフトウェアが注目されている。GIS の中でもオープンソースで無料で利用でき、高機能な QGIS は様々な分野で使用されている。

一方で、Web の情報をテキストマイニングし、共起ネットワークを作成することで、一見関係のない単語間の関係性を可視化する発想支援システムの研究が行われてきた [1]。両者ともにアプローチは違うものの様々な情報を可視化し、新たな知見を得るという点においては同じである。

つまり、GIS と発想支援システムを組み合わせることで両者を単独で使ったときよりも多くの知見や発想が得られると考えた。本研究では、QGIS と発想支援システムをシームレスに統合した QGIS プラグインの開発を目的とする。

キーワード：QGIS、スクレイピング、発想支援、プラグイン、ジオプロセシング

2 地理情報システムによる可視化

2.1 QGIS の主な活用事例

QGIS は、政府が公開しているオープンデータを使った商圏分析やコロナウイルスの分布を可視化するなど幅広い分野で使われている。また、多くのプラグインやオープンデータを使うことで、拡張性のある地理情報の分析ができる [2]。

東京都が公開しているコロナ感染者数をもとに市町村ごとのポリゴンで色分けすることで、東京都のコロナ感染者マップを可視化したものや、札幌市が公開しているヒグマ出没情報を利用し、ヒグマの出現場所を点で表しヒートマップで見やすくしたものなどがある (図 1 参照)。

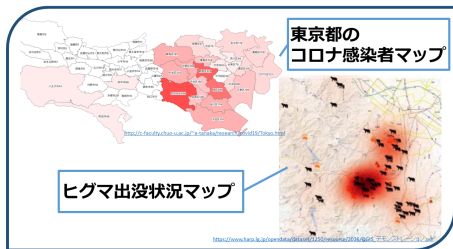


図 1：QGIS の活用事例

2.2 QGIS で扱えるデータ形式とプラグイン

QGIS は GIS データフォーマットの一つである Shape ファイルというものが使われる。Shape ファイルには道路や建物などの位置や形状、属性データをもつポイント、ライン、ポリゴンで構成されたベクタデータが格納されている。

Shape ファイルは複数のファイルから構成されている。主に、図形の情報を格納する shp ファイル、図形のインデックス情報を格納する shx ファイル、図形の属性情報をテーブルで格納する dbf ファイルの 3 つのファイルで構成

される。dbf ファイルは csv 形式で開くことができ、QGIS を使わなくてもファイルを編集することができる。

QGIS の機能の一つにプラグインの導入がある。プラグインとは拡張機能のことで、オープンソースのものを導入したり、自分で作成したものを導入することもできる。Googlemap の衛星画像などを QGIS 上に表示するプラグインや、位置情報が記載された画像の exif から Shape ファイルを作成し QGIS 上に表示するプラグインなどがある。

主に Python 言語で記述されており、プラグイン作成用のプラグインに Plugin Builder3 があり、プラグインの情報が記述された metadata.txt などの最低限必要なファイルを揃えてくれるものがある。(図 2 参照)

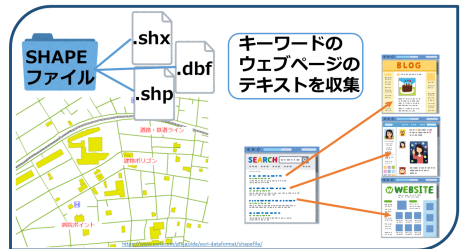


図 2：異なる情報を融合するプラグインの例

3 サイバー空間からの発想支援

3.1 サイバー空間からのテキストマイニング

現代社会においてインターネット上の情報は莫大になっており、今後も増え続けることが予想される。このインターネット上の情報を収集して QGIS で表示することができればより効率的にデータ分析することが可能になると思われる。

本研究では、キーワードごとに Google 検索から上から何件分かの URL を取得し、その URL からテキストを抽出した文章に対して HTML タグや JavaScript のコードを取り除くクリーニング処理をする。その次に、発想支援において必要な名詞や動詞などの形態素ごとに分解する形態素解析を行う [3]。

また、助詞の「し」や「が」助動詞の「です」は不要なので取り除く。そして単語群に対して正規化を行う。正規化することで文字種を統一できる。半角と全角や数字を統一することで同じ単語として扱うことができ、その単語群から共起ネットワークを作成するシステムが提案されている [1]。

3.2 発想支援の概要と手法

発想支援とは、人間の創造的問題解決における発想のプロセスを支援することである。この発想のプロセスでは、ある問題に対して、発散的思考・収束的思考・アイデア結晶化の 3 つのステップの流れで行われる。3 つのステップとはそれぞれ、関係のあると思われることがらを全て洗い出す 発散的思考と、それらの関係の整理・本質の追求を行う 収束的思考と、評価・決断を行うアイデア結晶化である。

図 2 は、「富山県立大学」というキーワードでテキストデータを収集し（発散的思考）、PageRank[4] で収集したデータを絞り込み（収束的思考）、3D 有向グラフを用いて発想支援を行った例である。線の太さが太いほどその単語同士に関係性があることを示している（図 3 参照）。

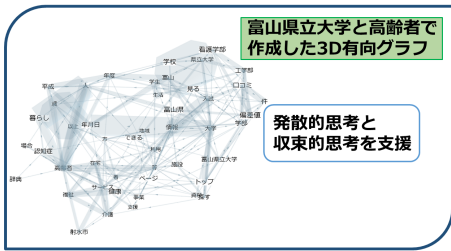


図 3：3D 有向グラフの例

4 提案手法

本研究では、QGIS でのオープンデータの可視化と発想支援に繋がる 3D 有向グラフを組み合わせ、それをシームレスで動作する QGIS プラグインの作成を目指す（図 4 参照）。

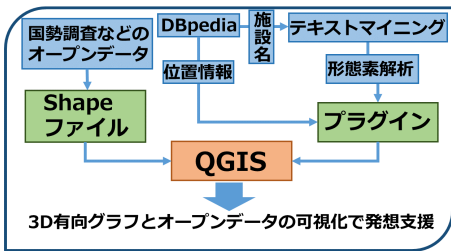


図 4：提案手法の概要図

QGIS で扱うオープンデータを組み合わせた Shape ファイルを作成するには、まず国土交通省が公開している行政区画データは支庁・振興局名などの余分な属性が含まれているので削除し、市町村区分の Shape ファイルを作成する。このとき、shp ファイルと dbf ファイルと shx ファイルが作成される。

次に、e-Stat 政府統計の総合窓口で公開されている国勢調査データなどを加工したものを先に作成した市町村区分の dbf ファイルと市町村名を基準に結合する。この Shape ファイルを使うことで、市町村別に人口や 65 歳以上の人口割合などのオープンデータの情報を色別グラデーションで表示することができる。

次に、QGIS プラグインの作成を行う。プラグインは Python で作成されている。学校などの施設の緯度経度情報や名称は DBpedia を使って抽出する。DBpedia から QGIS で選択されたレイヤーの都道府県の矩形範囲の緯度経度内にある施設を取得できるような SPARQL クエリを作成、発行し、JSON 形式で結果を取得する。

その情報を使ってベクタレイヤを作成することで QGIS 上に施設の点が表示される。その施設の点をクリックしたときにその施設の名称に対応する 3D 有向グラフが表示される。それを QGIS でのオープンデータを可視化したものと同時に見ることで発想支援につなげる。

DBpedia とは、Wikipedia の情報を抽出し、LOD(Linked Open Data) として公開しているコミュニティプロジェクトであり、Wikipedia 英語版を対象にしたものが DBpedia

である。また、Wikipedia 日本語版を対象にしたもので DBpedia japanese もある。この DBpedia を使う場合には、SPARQL(SPARQL Protocol and RDF Query Language) を使うことで DBpedia から必要な情報だけを取り出し利用することができる。

5 実験結果ならびに考察

QGIS 地図上に任意の表示させたいオープンデータの情報が色分けポリゴンで表示される。図 5 では、富山県の 65 歳以上の人口割合が赤ければ赤いほど多いように表示されている。また、富山県内の施設の一つを選択するとその施設名の 3D 有向グラフが右側に表示される。図 5 では、QGIS の地図上で富山県立大学が指定されており、それに合わせて富山県立大学と高齢者のキーワードでの 3D 有向グラフが右側に表示されている。

図 5 の 3D グラフから口コミ、認知症、暮らしの 3 つのワードと富山県の西部で高齢者の割合が高いことから、認知症の高齢者の近隣住民間の口コミを利用した見守りサービスを富山県西部で行うといった発想得られた。

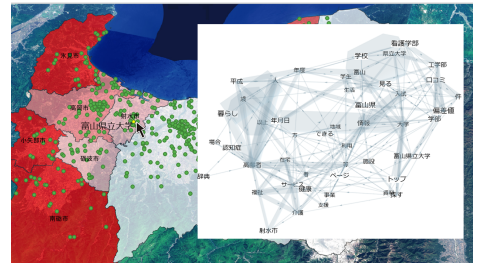


図 5：完成予想図

6 おわりに

本研究では、QGIS でシームレスに動作するようなプラグインを作成のための、QGIS で表示する Shape ファイルの作成とスクレイピングプログラムの調整ができた。現段階では、3D 有向グラフの表示と QGIS での表示が独立している。

今後の課題は、プラグインがシームレスに動作するようにプログラムを DBpedia からデータの取得するプログラムと 3D 有向グラフのプログラムをまとめることや、施設の点を選択した判定の処理方法、GUI をより見やすくより操作しやすいようなデザインにすることである。

参考文献

- [1] 山元悠貴, “Web 内容マイニングによる複数キーワードに対する 3D 有向グラフを用いた発想支援”, 富山県立大学学位論文 2019.
- [2] 谷川宮次, “GIS による「地域の知」の活用に関する一考察—オープンソース・オープンデータの活用を含めて—”, Bul. Hijiya Univ. Jun. Col., No. 54 2019.
- [3] 加藤耕太, “Python クローリング&スクレイピング—データ収集・解析のための実践開発ガイド”, 技術評論社 2017.
- [4] Rada Mihalcea, Paul Tarau, Elizabeth Figa, “PageRank on Semantic Networks with Application to Word Sense Disambiguation”, In Proceedings of the 20th International Conference on Computational Linguistics, pp. 1126–1132 2004.