

2025 年 1 月 10 日

「情報可視化の理論と実践」最終課題

教育学研究科 生涯学習基盤経営コース

博士課程 2 年 中尾康朗

(課題) 自身のとりくむ研究課題に関連して、問いを立てて、データ分析を行い、その問いを説明するような可視化作業を行う。

1. 答えたい問い

近年、図書館における情報サービスにおいて、資料や情報の調べ方に関する情報(調べ方情報あるいはパスファインダー情報、以下 PF)を提供するサービスが主なサービスのひとつとなっている。日本の各図書館が提供するサービスのうち、代表的なものに国立国会図書館のリサーチ・ナビ(RN)がある。 <https://ndlsearch.ndl.go.jp/rnavi>

RN では、特定のテーマに対して、調査のためにどのような情報源があるかをまとめている。それらの情報源には印刷体の情報源(参考図書)と、ネットワーク情報源(データベースや Web サイトなど)があり、個々のテーマを調べる際に利用できる有用なツール群である。

RN では、個々のテーマと各情報源はどのような関係にあるのだろうか、あるいは、個々のツールはどのような調べ方に有効なのだろうか。

- ・RN では実際、どのような情報源が掲載されているか

- ・各情報源は、どのような PF に結び付いているか

という点に着目して、調べ方情報と調査のための情報源の関係について可視化することを目標とした。

2. データの概要

2.1. データの用意

今回、分析対象としたのは、情報源のうち印刷体の情報源(参考図書、以下、参図)について確認を行った。これは、ネットワーク情報源(データベースや Web サイトなど)に関しては、RN の API(2.1.1.)から直接取得できるデータ中にリンク先情報が含まれておらず、これらを識別して取り出すためには別に工夫が必要なためである。

2.1.1. データの取得

データの取得方法は、RN の API を利用した。RN の API は「レファレンス協同データベース API2.0」仕様となっている。

https://crd.ndl.go.jp/jp/help/general/help_07_api_2.html

データの形式は XML 形式である。2024 年 12 月 30 日時点のデータ全件 1713 件を取り出した。

```
<result>4
<manual>4
<theme>官吏・軍人・公務員だった人物を探す</theme>4
<reg-id>ndl:s-rnavi-politics-public_officer</reg-id>4
<guide>国立国会図書館デジタルコレクション</guide>
国立国会図書館デジタルコレクションを利用した主題書誌はじめに近代日本の官
庁の歴史はすでに150余年を経過し、その間に、幾度もの変遷を重ねて今日に至っています。官庁機構の変遷およびそこに所属してきた職員（戦前は
官吏・軍人、戦後は公務員）の異動を把握するためには、官庁の職員録がもっとも基本的な資料です。当館は、所蔵する明治元年から昭和22年まで（
太政官制度時代から内務省解体まで）の期間に刊行された各種の職員録について、『職員録・職員録目録：明治元年～昭和22年 国立国会図書館所蔵
』国立国会図書館参考書誌部 編 1985年7月を刊行しました。現在では、当目録に掲載されている資料の多くが国立国会図書館デジタルコレクショ
ンで利用できるようになりました。このページは、当目録の掲載資料を基本とし、さらにその後に入れた資料やその一部に官吏・公務員の氏名情報
が掲載されている資料を追加して、国立国会図書館デジタルコレクションと紐づけで利用できるようにしたものです。現在でも各種の職員録が紙媒体
で刊行されており、当館が所蔵しているすべての資料がデジタルコレクションで利用できるわけではないことに留意ください。冊子資料・デジタル
化資料を使った官吏や公務員の人物情報の基本的な調べ方をリサーチ・ナビの以下のページでご案内しています。日本-公務員（官吏）の人物情報を
調べる（戦前編）日本-公務員の人物情報を調べる日本-官庁職員（公務員）の名簿中央官庁総合官員録 明治8(1875)年～明治18(1885)年官員録 明
治18(1885)年～明治28(1895)年『職員録』明治期（カラー版） 明治19(1886)年～明治45(1912)年『職員録』明治期（モノクロ版） 明治19(1886)年
～明治45(1912)年『職員録』大正期（カラー版） 大正2(1913)年～大正15(1926)年『職員録』大正期（モノクロ版） 大正2(1913)年～大正15(1926)
年『職員録』昭和前期（カラー版） 昭和2(1927)年～昭和18(1943)年『職員録』昭和前期（モノクロ版） 昭和2(1927)年～昭和18(1943)年『職員録』
昭和後期 昭和24(1949)年～昭和48(1973)年、昭和54(1979)年、昭和58(1983)年 ※『職員録』のモノクロ版、カラー版の原本は同一です。現在（2
024.12.26 現在）のところ、モノクロ版及び昭和後期は全文検索が可能です。参考情報国立公文書館所蔵の官員録・職員録 国立公文書館アジア歴史
資料センター「資料の閲覧・検索」アイコンをクリックし、次に「キーワード検索」をクリックして出てくる画面の左側にある資料詳細層表示を「国
立公文書館＞太政官・内閣関係＞官員録・職員録」とたどることで、国立公文書館が旧総理府から移管を受けた明治元年以来の戦前期の官員録・職員
録等のデジタル画像を年別に閲覧することができます。その他『主要職員録 加除式 改版』昭和17(1942)年12月1日現在『官廳・團體職員録：官
廳之部』昭和18(1943)年12月1日現在『公職追放に関する覚書該当者名簿』昭和23(1948)年2月刊 昭和24(1949)年2月刊 同左 追録付 同左資料
公職追放：『公職追放に関する覚書該当者名簿』復刻1 復刻2 官庁別明治期から昭和前期まで太政官内閣元老院枢密院開拓使宮内省警視庁外務
省内務省大蔵省陸軍省・陸軍省海軍省・海軍省司法省文部省文部省文部省工部省農商務省農林省商工省逓信省鉄道院 / 鉄道省厚生省拓務省企画院軍需省帝
会貴族院事務局帝会貴族院事務局昭和後期以降法務府 / 法務省 裁判所外務省大蔵省文部省農林省 / 農林水産省電気通信省郵政省厚生省建設省
運輸省労働省自治省防衛庁警察庁中央諸機関中央気象台帝国国士院 / 日本学士院旧外地等台湾総督府朝鮮総督府樺太庁関東庁南洋庁参考 満洲（国）
昭和前期までの教育機関・その他帝国大学旧制高等学校ビジネス系学校：高等商業学校・商科大学 商船学校 外国語学校技術系学校農林水産系学校教
員養成学校医療系学校芸術系・神道系学校 音楽学校 神宮皇学館盲学校・聾啞学校軍人養成学校御雇外国人地方官庁 / 地方自治体総合官員
録 明治8(1875)年～明治18(1885)年官員録 明治18(1885)年～明治28(1895)年『職員録』明治期（カラー版） 明治19(1886)年～明治45(1912)年『
職員録』明治期（モノクロ版） 明治19(1886)年～明治45(1912)年『職員録』大正期（カラー版） 大正2(1913)年～大正15(1926)年『職員録』大正
期（モノクロ版） 大正2(1913)年～大正15(1926)年『職員録』昭和前期（カラー版） 昭和2(1927)年～昭和18(1943)年『職員録』昭和前期（モノク
ロ版） 昭和2(1927)年～昭和18(1943)年『職員録』昭和後期 昭和24(1949)年～昭和45(1970)年 昭和47(1972)年※モノクロ版、カラー版ともに原本
は同一です。現在（2024.8.4 現在）のところ、全文検索ができるのはモノクロ版のみです。都道府県別参考 府県史料（国立公文書館デジタルアーカ
イブ） 明治期以来の地方官庁・地方自治体 職員録のほか、その地域の年鑑類等でも人名録が掲載されているもの。旧県北海道青森岩手宮城秋田山形
福島茨城栃木群馬埼玉千葉東京神奈川新潟富山石川福井山梨長野岐阜静岡愛知三重滋賀京都大阪兵庫奈良和歌山鳥取島根岡山広島山口徳島香川愛媛高
知福岡佐賀長崎熊本大分宮崎鹿児島沖縄地方官庁昭和前期までの中学校・高等女学校</guide>4
<completion>1</completion>4
<note>リサーチ・ナビで見る：<a href="https://ndlsearch.ndl.go.jp/rnavi/politics/public_officer">https://ndlsearch.ndl.go.jp/rnavi/politics/public_officer</a></note>4
<system>4
<reg-date>20231027143346</reg-date>4
<lst-date>20241226180000</lst-date>4
<sys-id>2000030369</sys-id>4
<lib-id>1110001</lib-id>4
<lib-name>国立国会図書館 (National Diet Library) </lib-name>4
<file-num>0</file-num>4
</system>4
<url><a href="https://crd.ndl.go.jp/reference/detail?page=man_view&id=2000030369">https://crd.ndl.go.jp/reference/detail?page=man_view&id=2000030369</a></url>4
</manual>4
</result>4
```

図表 1 RN データ例

2.1.2. データの前処理

ライブラリ (Beautiful Soup) を利用し、必要なタグのデータのみ抽出 (theme, note, guide, reg-id) した。guide 中の自由記述文の中に参図が記載されている。正規表現を使
って抽出した。テーマ 1,713 件、参図 14,218 件であった。

調査テーマ	theme
管理番号	reg-id
調べ方	guide
備考	note

図表 2 RN 抽出項目名

その後、雑然データから整然データへの加工を行った。1つの guide の中に複数の参図の情報源が含まれている雑然データの状態から、整然データの条件(1つの列は1つの変数を表す。1つの行は1つの観測を表す。1つの表は1つの観測単位を持つ。1つのセルは1つの値を表す。)を満たすデータにフォーマットを変換した。そして、1つの PF の中に、同じ参図がくりかえし登場する場合はユニークにした。なお、各 PF には利用する際の注意事項として『資料名』と記載されている箇所があり、これらは参図ではないため前処理の段階で除外した。

3. データ変数と視覚変数の対応関係

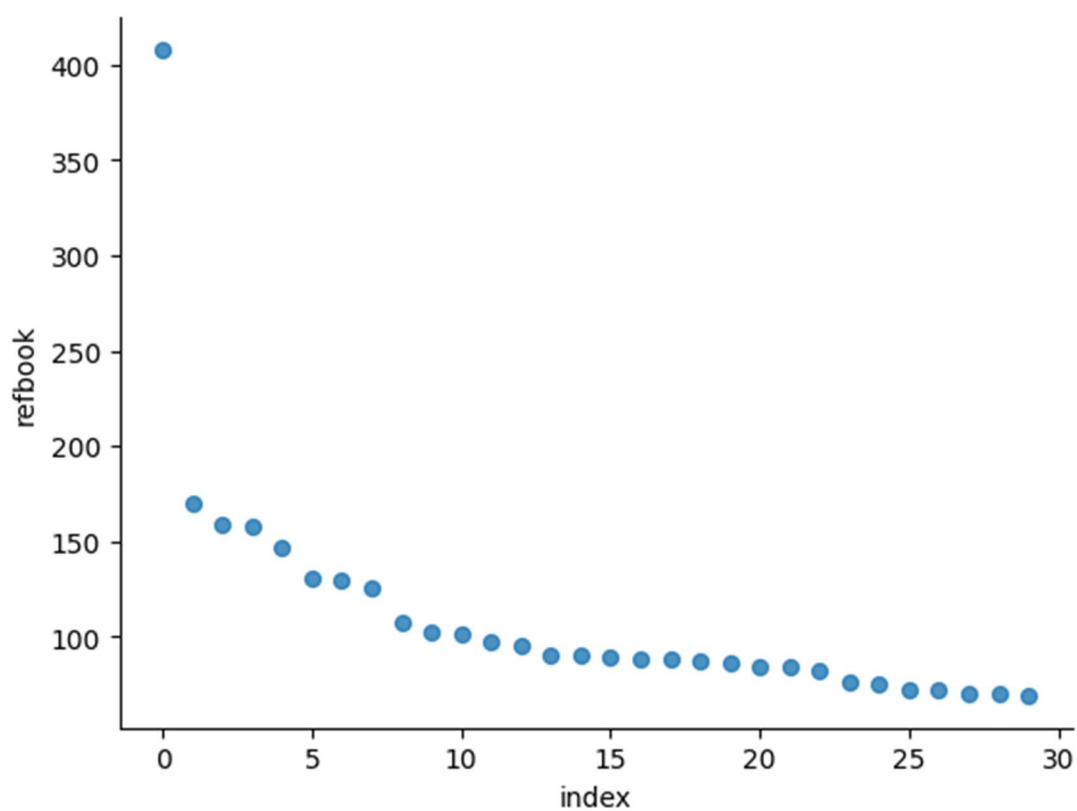
3.1. どのような参考図書が掲載されているか

3.1.1. 各 PF に含まれる参図

各 PF に含まれる参図について集計した。(PG1)

index	theme	refbook
0	日本-官庁の年史	408
1	北海道の写真集	170
2	東京都の写真集(都全域・広域、区部)	159
3	神奈川県の写真集	158
4	新潟県の写真集	147
5	兵庫県の写真集	130
6	埼玉県の写真集	129
7	政治分野のおもな参考図書	125
8	大衆演芸について調べる	107
9	長野県の写真集	102
10	新聞社の社史を調べる(日本)	101
11	大阪府の写真集	97
12	明治・大正時代の主な新聞とその参考文献(関東地域)	95
13	茨城県の写真集	90
14	富山県の写真集	90
15	日本の地名を調べる	89
16	岩手県の写真集	88
17	録音資料の出版目録(ポピュラー音楽)	88
18	福島県の写真集	87
19	地方紙の記事索引・検索サービス	86
20	中国の「正史」の日本語訳	84
21	長崎県の写真集	84
22	哲学・思想に関する文献を探すには(主題書誌)	82
23	岐阜県の写真集	76
24	宮城県の写真集	75

図表 3 各 PF 参図頻度(上位 25)



データ変数	データ変数の性質	視覚変数	視覚変数の性質
テーマ	順序尺度	位置 X	定量性・順序性
参図	比例尺度	位置 Y	定量性・順序性

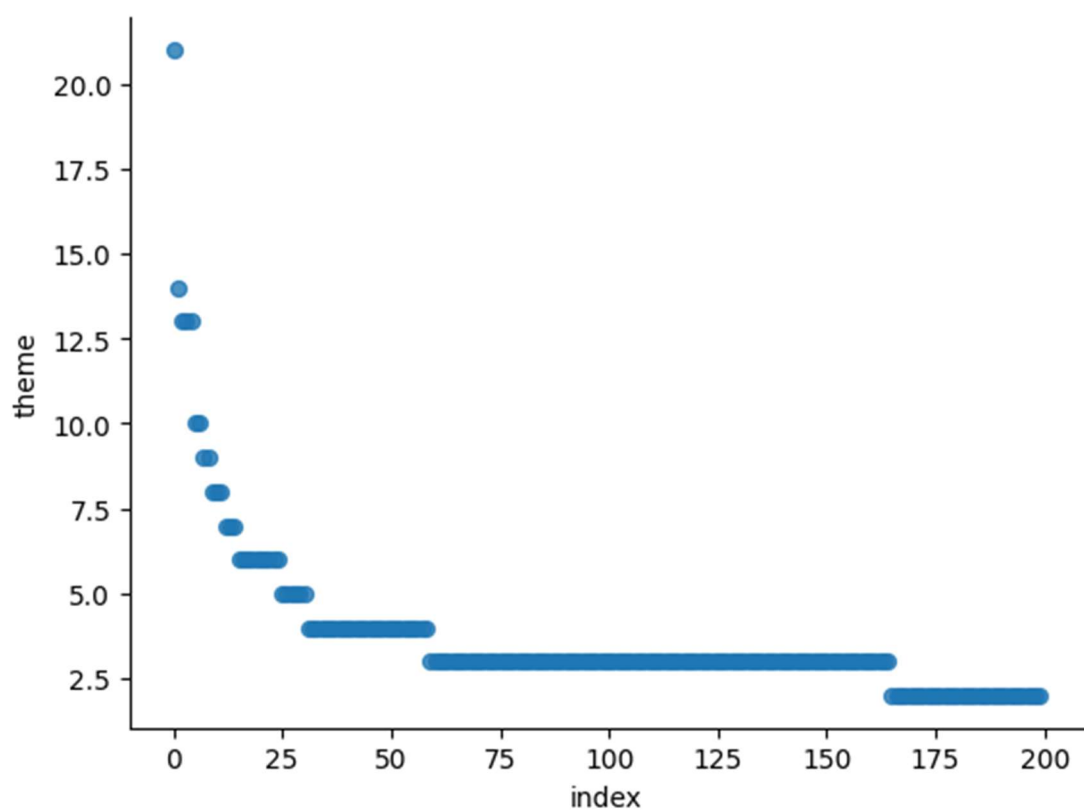
図表 4 データ：参図

3.1.2. 各参図を含む PF

各参図を含む PF について集計した。(PG2)

index	refbook	theme ▼
0	『参考書誌研究』	21
1	『写真で見る日本』	14
2	『レファレンス』	13
3	『国立国会図書館月報』	13
4	『官報』	13
5	『職員録』	10
6	『雑誌新聞総かたろぐ』	10
7	『中国商工地図集成』	9
8	『出版年鑑』	9
9	『近代アジア・アフリカ都市地図集成』	8
10	『日本新聞年鑑』	8
11	『近代中国都市地図集成』	8
12	『経済構造実態調査』	7
13	『国史大辞典』	7
14	『出版指標年報』	7
15	『運輸関係団体名簿』	6
16	『運輸関係団体等名簿』	6
17	『歴史読本』	6
18	『週刊東洋経済』	6
19	『レジャー白書』	6
20	『新風土記：1954-1958』	6
21	『角川日本地名大辞典』	6
22	『最新運輸業界の動向とカラクリがよくわかる本』	6
23	『日本歴史地名大系』	6
24	『日本の道』	6

図表 5 各参図 PF 頻度（上位 25）



データ変数	データ変数の性質	視覚変数	視覚変数の性質
参図	順序尺度	位置 X	定量性・順序性
テーマ	比例尺度	位置 Y	定量性・順序性

図表 6 データ：PF

3.1.3. 参図に含まれるワード

RN で採用されている参図のワードの傾向を探索的に分析した。(PG3)

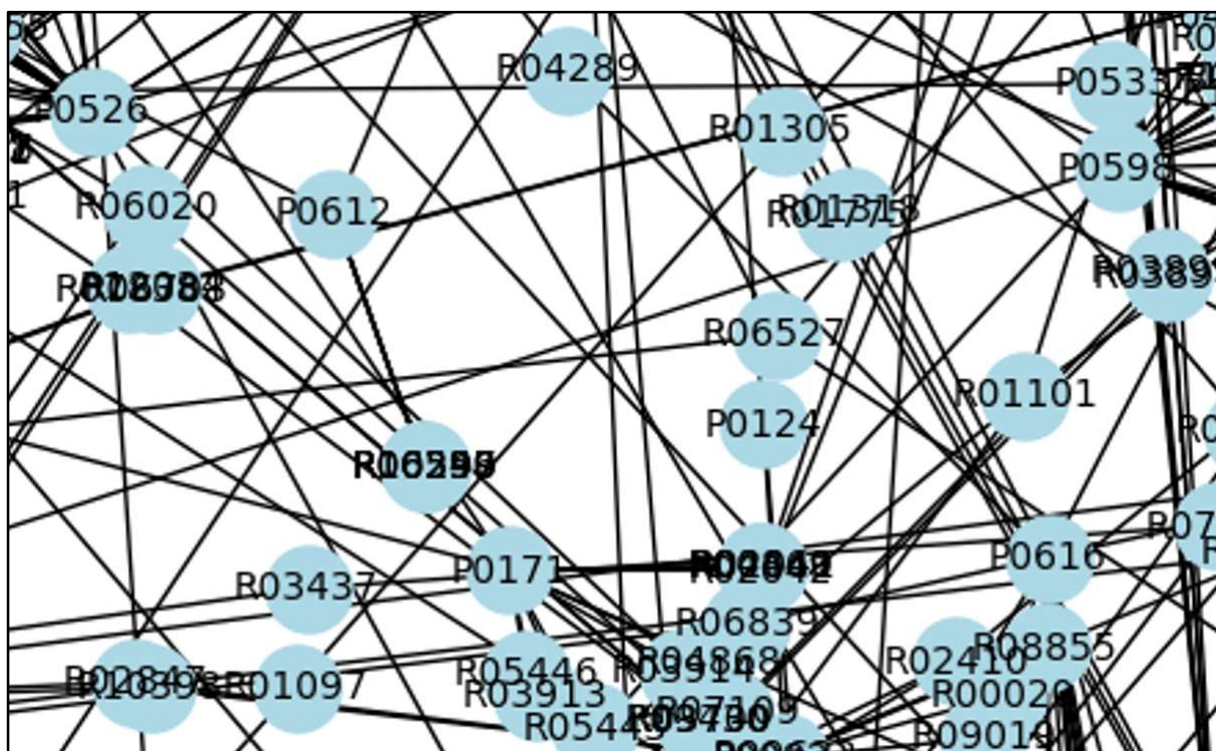
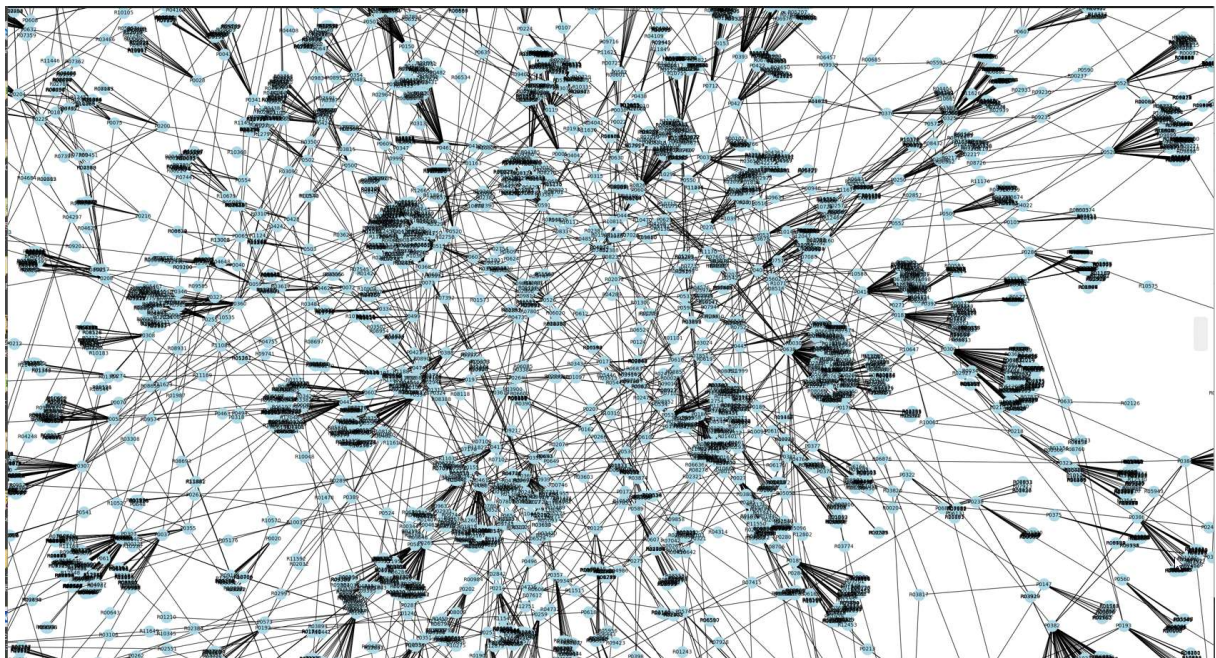


データ変数	データ変数の性質	視覚変数	視覚変数の性質
ワード位置	テキスト	位置	選択性
ワード大きさ	テキスト	大きさ	定量性・順序性
ワード色	テキスト	色	選択性

図表 7 データ：ワード（参図）

3.2. 各参図はどのような PF に結び付いているか

各参図がそれぞれの PF と結び付いているか、ネットワークを分析した。日本語のテーマと参図だと見つらくなるため、PF は P0001, 参図は R00001 のようにコード化して処理を行った。(PG4)



データ変数	データ変数の性質	視覚変数	視覚変数の性質
テーマ	テキスト	位置	選択性
テーマ	テキスト	大きさ	選択性
テーマ	テキスト	色	選択性

データ変数	データ変数の性質	視覚変数	視覚変数の性質
テーマ	テキスト	位置	選択性
テーマ	テキスト	大きさ	選択性
テーマ	テキスト	色	選択性
リンク	名義尺度	線	関連性

図表 8 データ：PF⇔参図

4. 可視化作品と考察

3.1.1.や 3.1.2.では、両者ともロングテール的な分布になっていることが確認できる。散布図を使ったが、棒グラフでも同じ効果が得られると思われた。

3.1.3.は、探索的な可視化として、日本語形態素解析を行って、参図のタイトルから名詞のみ取り出し、ワードクラウドにしてみた。ワードクラウドの場合、大きさは頻度を表しているものの、色や位置はランダムで、処理される都度変化するということもあり、ワードにある程度、偏りがありそうだという傾向は感じられるが、特にはっきりとした言葉や分野上の特徴は見い出せなかった。

3.2 では、ネットワーク図を使って、PF と参図の関係性をつかもうと試みた。可視化はできたものの、ノードが密集している箇所は拡大しても確認できなかった。ライブラリやパラメータなどを調整することで解決できるかもしれないが、うまくコントロールできるところまでコーディングを理解することができなかった。また、リソースを使うためか 1 回の処理に時間がかかり、その点も難しく感じた。また、ネットワーク図のような場合、データ変数と視覚変数の関係をどのように整理したらよいか、という点もうまく判断できなかった。

5. コード

(PG1)/(PG2)

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
from matplotlib import font_manager

# CSV ファイルを読み込む
file_path = "RN_refbooks_uniq.csv"
df_plot1 =
df.value_counts("theme").reset_index(name="refbook").head(733)

df_plot1

df_plot2 =
df.value_counts("refbook").reset_index(name="theme").head(14218)
```


df_plot2

```
from matplotlib import pyplot as plt
_df_9.plot(kind='scatter', x='index', y='theme', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)

from matplotlib import pyplot as plt
_df_2.plot(kind='scatter', x='index', y='refbook', s=32, alpha=.8)
plt.gca().spines[['top', 'right',]].set_visible(False)
```

(PG3)

```
import csv
import collections
from janome.tokenizer import Tokenizer
from wordcloud import WordCloud

list = []
with open('RN_refbooks_uniq.csv', 'r') as f:

    # カラムの値を抽出
    for row in csv.reader(f):
        list.append(row[3])

    # 先頭文字を削除
    del list[0]

    # 文字列をつなげる
    b = ""
    for a in reversed(list):
        b += a

    # 文字の整形(改行削除)
    text = "".join(b.splitlines())

    # 単語ごとに抽出
    docs=[]
    t = Tokenizer()
    tokens = t.tokenize(text)
    for token in tokens:
        if len(token.base_form) > 2:
            docs.append(token.surface)

    ## wordcloud の実行
    ## 日本語フォントを指定
    c_word = ' '.join(docs)
    # wordcloud = WordCloud(background_color='white',
    #
    font_path='/usr/share/fonts/opentype/ipaexfont-gothic/ipaexg.ttf',
    #
    width=800, height=400).generate(c_word)

    wordcloud = WordCloud(font_path =
'/usr/share/fonts/truetype/fonts-japanese-mincho.ttf',
background_color="white",
```

```

width=1000,height=400).generate(text)

## 結果を画像に保存
wordcloud.to_file('wordcloud.png')

# 単語を多い順に並べる
c = collections.Counter(docs)
print(c)

# プロット
plt.figure(figsize=(8, 4))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.tight_layout(pad=0)
plt.savefig("/content/drive/MyDrive/Colab
Notebooks/wordcloud.png", dpi=300)
plt.show()

```

(PG4)

```

import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
from matplotlib import font_manager

# CSV ファイルを読み込む
file_path = "pr_pair.csv"
data = pd.read_csv(file_path, header=None)

# 1 列目と 2 列目のデータを使用してエッジを作成
edges = list(zip(data[0], data[1]))

# ネットワークグラフを作成
G = nx.Graph()
G.add_edges_from(edges)

# エッジをハイライトする設定 (例: 最初のエッジをハイライト)
highlight_edges = [edges[0]]
edge_colors = ["red" if edge in highlight_edges else "black" for edge
in G.edges()]
edge_widths = [2.5 if edge in highlight_edges else 1 for edge in
G.edges()]

# ネットワーク図を描画
plt.figure(figsize=(200, 200))
pos = nx.spring_layout(G, seed=42)
nx.draw(
    G, pos, with_labels=True,
    labels={node: node for node in G.nodes()},
    node_color='lightblue', node_size=500, font_size=10,
    #font_family=jp_font.get_name(),
    edge_color=edge_colors, width=edge_widths
)
plt.title("Network Graph with Highlighted Edges")

```

```
plt.show("pr_pair.png")
```