

# テキストマイニング

## — Part 4 —

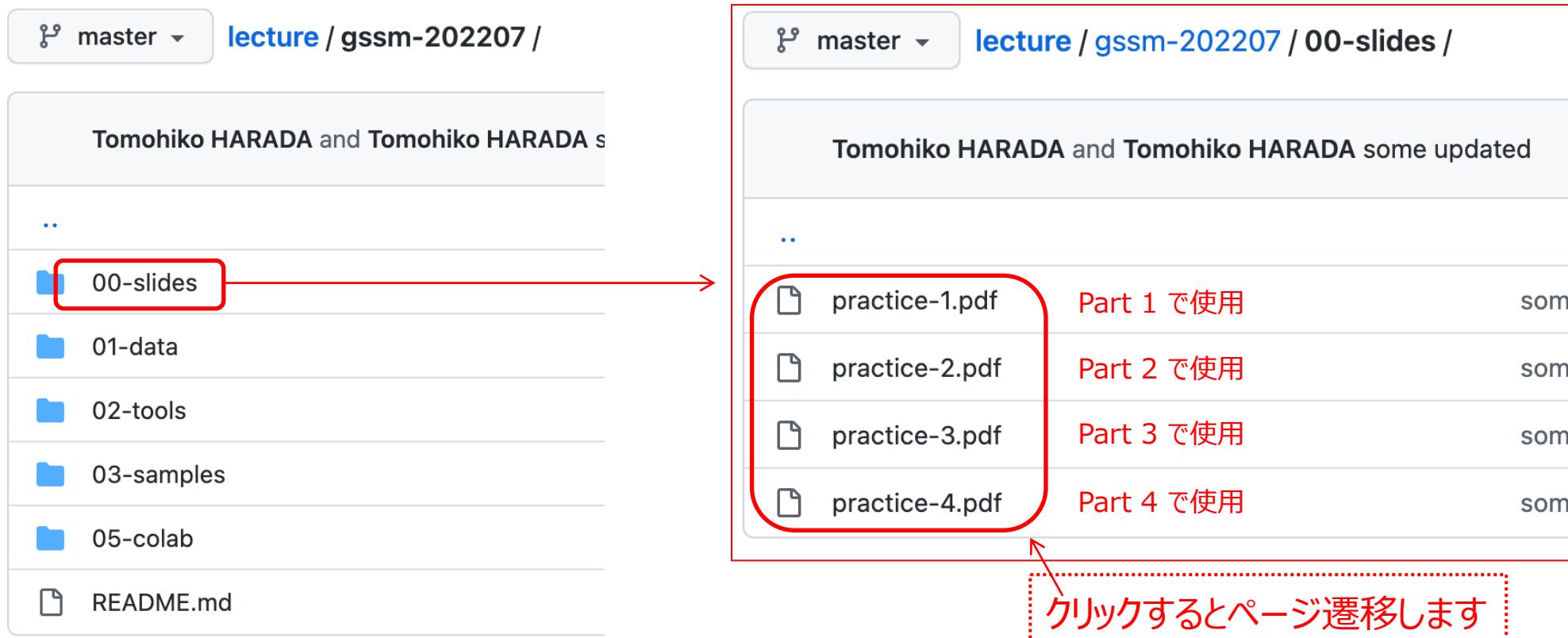
R04年度  
人文社会ビジネス科学学術院  
ビジネス科学研究群

# スケジュール

- Part 1
  - 説明 — 自然言語処理のトレンド
  - 説明 — 環境説明
- Part 2
  - 説明 — テキストマイニングの手順
  - 説明 — データ理解
  - 実習 — データ理解 (Excel)
- Part 3
  - 説明 — データ分析の実践 (KHCoder)
  - 実習 — データ分析の実践 (KHCoder)
- Part 4
  - 実習 — データ分析の実践 (KHCoder)
  - 実習 — データ分析の実践 (KHCoder)

# 講義スライド

- <https://github.com/haradatm/lecture/tree/master/gssm-202207>



# (復習) テキストマイニングの手順

## ・データをよく知る

- ・データ件数や構成比を集計 → データを理解する
  - ・旅行目的別の人気エリアは?
  - ・同伴者別の人気エリアは?
  - ・数値評価による人気エリアの差異は?

## ・テーマを設定する

- ・解決すべき課題を決める → 分析目的を明確にする
  - ・数値評価が低い原因は?
  - ・高評価の施設に学ぶ改善点は?

## ・データ分析に取り組む

- ・これら課題を解決するために、テキスト分析を実施

# (復習) クチコミサイトの例



- ホテルのクチコミ数: 1,237万件 ※年間約60~70万

The screenshot shows the Rakuten Travel website at <https://travel.rakuten.co.jp/review/>. The main heading is 'お客様の声' (Customer Reviews) with a count of 'ホテルのクチコミ数No.1 12,369,840'. Below this, there's a search bar for reviews and a section for new reviews. On the right, there's a summary box stating: '「お客様の声」には、実際にご利用になった方のご意見・ご感想が満載です。' (There are many opinions and feelings from customers who actually used the service). The page also features a green navigation bar with tabs for '国内' (Domestic) and '海外' (Overseas).

経年変化:

780万件 (2015)  
→ 836万件 (2016)  
→ 900万件 (2017)  
→ 973万件 (2018)  
→ 1,042万件 (2019)  
→ 1,098万件 (2020)  
→ 1,165万件 (2021)  
→ **1,237万件 (今回)**  
※ 2021/6/4現在

## 鴨川シーワールドホテルのクチコミ・お客様の声

[●ホテル・旅行のクチコミTOPへ](#)

## 総合評価

4.12

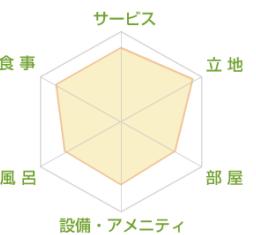
アンケート件数：886件

## 評価内訳

- 5点 ■■■■■ 236件
- 4点 ■■■■ 302件
- 3点 ■■ 47件
- 2点 ■ 15件
- 1点 ■ 9件

## 項目別の評価

サービス	4.11
立地	4.61
部屋	3.53
設備・アメニティ	3.62
風呂	3.53
食事	4.10



## 総合 2

## 投稿者さんの 鴨川シーワールドホテル のクチコミ（感想）



投稿者さん

2015年06月11日 17:03:57

良かったところ

- ・部屋からの景色（朝日最高でした）
- ・食事（品数多く、朝夕とも良かったです）
- ・フロントの方の対応（お姉さんがとても頑張っていました）以上。

掃除が行き届いているとの口コミを多く見ましたが、それは思ひませんでした。

気にかかることは多々ありましたが、フロントのお姉さんが一生懸命で、その笑顔に救われた思いです。

## 評価

... 総合 2

サービス 2

立地 4

部屋 4

設備・アメニティ 2

風呂 2

食事 4

## 旅行の目的

... レジャー

## 同伴者

... 家族

## 宿泊年月

... 2015年06月

## 情報



鴨川シーワールドホテル

2015年06月11日 19:32:50

この度は、ご利用頂きまして誠にありがとうございます。

客室内清掃の件、大変申し訳ござい

重要改善として、早急に対応いたします。

今後は、この様な事の無いように、清掃・点検を強化いたします。

フロントスタッフへのお言葉  
誠にありがとうございます。

モチベーションアップに繋がる  
お客様からの声として、  
スタッフと共有させて頂きます。

機会がございましたら、またご利用をお待ちしております。

## テキストデータ

## 数値評価

# (再掲) 実習で使用するデータ

## 楽天トラベル のクチコミデータ

- ・収集期間は **2019-2020** および **2021-2022(～GW明け)** の **2セット**
- ・以下の **10 エリアごと** 同数に **1,000件ずつ** ランダムサンプリング
- ・データ件数は **1万件** × 2セット

レジャー	5エリア	登別, 草津, 箱根, 道後, 湯布院	<b>1,000件</b> × 10エリア = 計10,000件
ビジネス	5エリア	札幌, 名古屋, 東京, 大阪, 福岡	

# (再掲) 実習で使用するデータ

## 楽天トラベル のクチコミデータ

- データ項目は **18項目** (テキスト1項目+その他の属性**17項目**)

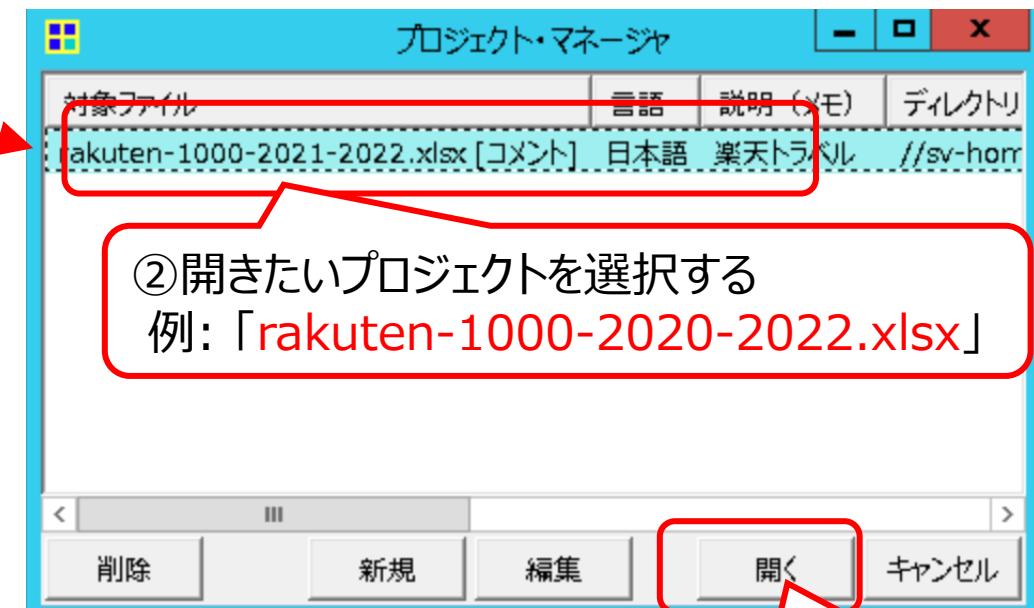
施設情報	4項目	カテゴリ, エリア, 施設番号, 施設名
口コミ	1項目	コメント (テキスト)
ユーザー評価	7項目	総合, サービス, 立地, 部屋, 設備・アメニティ, 風呂, 食事
その他の分類	2項目	旅行の目的, 同伴者
宿泊日	1項目	宿泊年月
ユーザー情報	3項目	ユーザー, 年代, 性別

# 準備 — 作成済みのプロジェクトを開く

①メニューから「プロジェクト」「開く」を選択 (注)



注: 次回 KH Coderを起動した時は「新規」ではなく  
「開く」を選択します



②開きたいプロジェクトを選択する  
例:「rakuten-1000-2020-2022.xlsx」

⑤「開く」をクリック

# 数値評価で違いを見るのは難しい

- ユーザーの8割が4~5の評価,  
1~2をつけない→本音が見えない

数値評価の平均 (エリア別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.29	4.29	4.18	4.07	4.34	4.29	4.34
01_登別	4.08	4.20	3.96	3.87	4.33	4.13	4.17
02_草津	4.29	4.27	4.13	4.04	4.38	4.18	4.33
03_箱根	4.26	4.16	4.18	4.05	4.28	4.25	4.27
04_道後	4.26	4.42	4.21	4.05	4.28	4.25	4.36
05_湯布院	4.58	4.39	4.40	4.05	4.28	4.25	4.58
B_ビジネス	4.14	4.40	4.22	4.05	3.94	4.32	4.32
06_札幌	4.17	4.42	4.26	4.07	3.96	4.15	4.35
07_名古屋	4.07	4.29	4.17	3.99	3.91	4.03	4.24
08_東京	4.13	4.43	4.20	4.04	3.88	4.21	4.32
09_大阪	4.16	4.42	4.20	4.04	3.88	4.17	4.37
10_福岡	4.17	4.43	4.20	4.04	3.94	4.25	4.32

数値評価の平均 (レジャー, ビジネス別)

行ラベル	平均 / サービス	平均 / 立地	平均 / 部屋	平均 / 設備・アメニ	平均 / 風呂	平均 / 食事	平均 / 総合
A_レジャー	4.29	4.29	4.18	4.07	4.34	4.29	4.34
B_ビジネス	4.14	4.40	4.22	4.05	3.94	4.16	4.32

# 実践編 — 特徴語の集計

- 宿泊客は、どの項目に注目しているか?
  1. カテゴリー「レジャー」と「ビジネス」を比較する
  2. カテゴリー「レジャー」(or「ビジネス」) の 5エリアを比較する

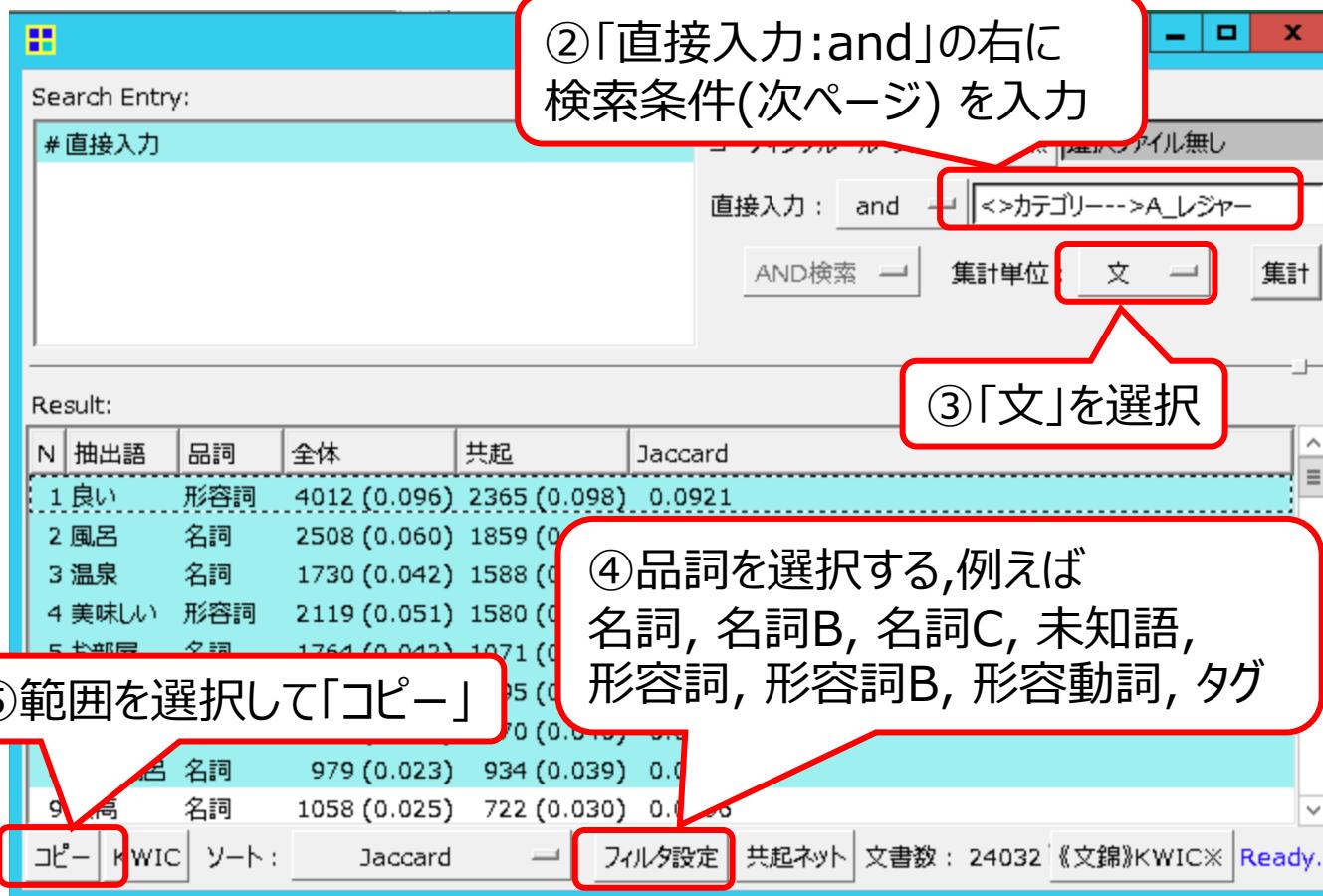
- 手順
  - テキスト中の特徴語を集計

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]”<>カテゴリ--->A\_レジャー”」「集計単位:文」→「フィルタ設定」→「品詞=名詞, 形容動詞, 未知語, タグ, 形容詞, 名詞B, 形容詞B, 名詞C」を選択→「集計」→結果を選択し「コピー」

- エリアによって特徴語がどう異なるかを比較
- 注目する項目の違いを考察する

# 実践編 — 特徴語の集計

- ①メニューから「ツール」「抽出後」「関連語検索」を選ぶ



- ⑥EXCEL にペースト

A1	B	C	D	E	F
1	1 良い	形容詞	4012 (0.096)	2365 (0.098)	0.0921
2	2 風呂	名詞	2508 (0.060)	1859 (0.060)	0.0753
3	3 温泉	名詞	1730 (0.042)	1588 (0.042)	0.0657
4	4 美味しい	形容詞	2119 (0.051)	1580 (0.051)	0.0643
5	5 お部屋	名詞	1764 (0.043)	1071 (0.043)	0.0433
6	6 スタッフ	名詞	1588 (0.043)	995 (0.043)	0.0404
7	7 宿	名詞C	1046 (0.023)	970 (0.040)	0.0402
8	8 露天風呂	名詞	979 (0.023)	934 (0.039)	0.0388
9					

# 直接入力: [and] の右側に入力する条件

レジャー:

<>カテゴリ-->A\_レジャー

<>エリア-->01\_登別

<>エリア-->02\_草津

<>エリア-->03\_箱根

<>エリア-->04\_道後

<>エリア-->05\_湯布院

ビジネス:

<>カテゴリ-->B\_ビジネス

<>エリア-->06\_札幌

<>エリア-->07\_名古屋

<>エリア-->08\_東京

<>エリア-->09\_大阪

<>エリア-->10\_福岡

# 使い方 – 外部変数(エリア)を利用する

①メニューから「ツール」「外部変数と見出し」を開く



A	B	C	D	E	F	G	H	I	J	K
1										
2	01_登別		02_草津		03_箱根		04_道後			
3	食事	.059	温泉	.068	思う	.066	温泉	.054		
4	良い	.058	湯畑	.064	食事	.064	良い	.051		
5	風呂	.057	風呂	.062	良い	.060	朝食	.045		
6	思う	.054	良い	.061	風呂	.053	ホテル	.042		
7	温泉	.049	食事	.056	美味しい	.049	美味しい	.042		
8	美味しい	.044	草津	.055	露天風呂	.048	道後	.041		
9	宿泊	.043	満足	.042	お部屋	.045	対応	.028		
10	満足	.041	美味しい	.042	温泉	.043	松山	.028		
11	料理	.033	宿	.041	満足	.043	立地	.026		
12	行く	.032	行く	.037	料理	.034	大変	.023		
13	05_湯布院		06_札幌		07_名古屋		08_東京			
14	食事	.072	ホテル	.061	ホテル	.063	利用	.060		
15	美味しい	.062	部屋	.058	名古屋	.059	部屋	.057		
16	宿	.061	朝食	.057	朝食	.058	ホテル	.054		
17	風呂	.059	利用	.055	利用	.055	宿泊	.039		
18	露天風呂	.050	札幌	.055	部屋	.055	朝食	.035		
19	料理	.049	良い	.052	思う	.047	快適	.034		
20	満足	.048	宿泊	.043	フロント	.035	お部屋	.034		
21	宿泊	.044	対応	.034	綺麗	.032	駅	.034		
22	温泉	.043	広い	.033	駅	.030	立地	.034		
23	お部屋	.042	立地	.031	対応	.029	フロント	.032		
24	09_大阪		10_福岡							
25	ホテル	.061	ホテル	.060						
26	利用	.056	利用	.060						
27	部屋	.050	部屋	.058						
28	宿泊	.040	朝食	.040						
29	立地	.039	博多	.039						
30	朝食	.039	立地	.039						
31	駅	.036	宿泊	.036						
32	綺麗	.033	便利	.031						
33	便利	.031	広い	.033						
34	フロント	.030	駅	.030						

各エリアの特徴語を10件ずつ  
一覧 (数値は Jaccard係数)

# 実践 — 特徴語の集計例

- 数値評価ではすべての項目に回答  
→ レジヤーとビジネスでは注目する項目にかなり偏りがありそう

A_レジヤー	数値評価指標
良い	.092
風呂	.075
温泉	.066
美味しい	.064
お部屋	.043
スタッフ	.040
宿	.040
露天風呂	.039
最高	.030
夕食	.029

数値評価指標
風呂
部屋
食事
サービス
設備
立地

01_登別	02_草津	03_箱根	04_道後	05_湯布院					
良い	.058	温泉	.068	良い	.060	温泉	.054	美味しい	.062
風呂	.057	湯畑	.064	風呂	.053	良い	.051	宿	.061
温泉	.049	風呂	.062	美味しい	.049	ホテル	.042	風呂	.059
美味しい	.044	良い	.061	露天風呂	.048	美味しい	.042	露天風呂	.050
ない	.037	美味しい	.042	お部屋	.045	立地	.026	温泉	.043
スタッフ	.031	宿	.041	温泉	.043	よい	.025	お部屋	.042
バイキング	.030	ない	.036	スタッフ	.034	大変	.023	スタッフ	.038
夕食	.028	最高	.031	宿	.033	浴場	.022	最高	.031
残念	.027	スタッフ	.030	夕食	.030	残念	.022	家族	.027
最高	.025	露天風呂	.028	残念	.024	夕食	.021	大変	.026

B_ビジネス	数値評価指標
部屋	.105
ホテル	.095
立地	.045
ない	.044
広い	.038
綺麗	.038
便利	.038
フロント	.037
駅	.036
快適	.034

数値評価指標
風呂
部屋
食事
サービス
設備
立地

06_札幌	07_名古屋	08_東京	09_大阪	10_福岡					
ホテル	.061	ホテル	.063	部屋	.057	ホテル	.061	ホテル	.060
部屋	.058	部屋	.055	ホテル	.054	部屋	.050	部屋	.058
良い	.052	フロント	.035	快適	.034	立地	.039	立地	.041
広い	.033	綺麗	.032	お部屋	.034	駅	.036	便利	.034
ない	.033	駅	.030	駅	.034	綺麗	.033	広い	.032
立地	.031	便利	.029	立地	.034	ない	.031	駅	.030
便利	.031	立地	.029	ない	.033	便利	.031	綺麗	.029
綺麗	.030	快適	.027	フロント	.032	フロント	.030	近い	.028
フロント	.030	浴場	.024	便利	.031	広い	.030	フロント	.026
駅	.029	近い	.022	近い	.027	近い	.028	大変	.026

Tips: 「ツール」→「外部変数と見出し」→「リスト」→「変数リスト=カテゴリー」を選択→「▽特徴語」→「選択した値」→「関連語検索画面」→「フィルタ設定」→「品詞=名詞、形容動詞、未知語、タグ、形容詞、名詞B、形容詞B、名詞C」を選択→「▽特徴語」→「一覧(EXCEL形式)」で連続実行

# Tips: 表記ゆれを吸収する (1/3)

出所: <https://github.com/ko-ichi-h/khcoder/issues/101>

- 目的

- 同じ意味の単語を同一視する別の単語として扱わない  
例) 「部屋」「お部屋」の 2単語 → どちらも「部屋」としてカウント

- 方法

- 「表記揺れを吸収」プラグインを利用する

- 手順

1. プラグインをダウンロードし, 解凍して **plugin\_jp** 配下へコピー

[ダウンロード URL] [https://github.com/ko-ichi-h/khcoder/files/4809463/z1\\_edit\\_words3.zip](https://github.com/ko-ichi-h/khcoder/files/4809463/z1_edit_words3.zip)

[解凍後ファイル名] z1\_edit\_words3.zip → z1\_edit\_words3.pm

[配置後のパス] khcoder3¥**plugin\_jp**¥z1\_edit\_words3.pm

(次ページにつづく)

# Tips: 表記ゆれを吸収する (2/3)

- 手順

- 手順
- プラグインファイル

**z1\_edit\_words3.pm** を編集する

```
1 package z1_edit_words3;
2 use utf8;
3
4 my $config = {
5     '友達' =>
6         [
7             '友人',
8             '旧友',
9             '親友',
10            '盟友',
11            '友',
12        ],
13        '格別' =>
14        [
15            '特別',
16            '格別', # 通常
17        ], # の
18        '偶然' =>
19        [
20            '偶然', # 形容
21        ],
22    };
23 }
```



```
1 package z1_edit_words3;
2 use utf8;
3
4 my $config = {
5     '部屋' =>
6         [
7             'お部屋',
8         ],
9 };
```

編集前

編集後

- ↓
- KH Coder を再起動する
  - プロジェクトファイルを開く
  - メニューから「ツール」「プラグイン」「表記ゆれの吸収」を選ぶ
  - 分析を続ける

適用後の例 →

「部屋」と「お部屋」が  
ひとつの単語にまと  
まっている

#	抽出語	品詞/活用	頻度
1	部屋	名詞	6737
2	お部屋	名詞	4876
3	大部屋	名詞	1861

# Tips: 表記ゆれを吸収する(3/3)

A_レジヤー	数値評価指標
良い	.092
風呂	.075
温泉	.066
美味しい	.064
スタッフ	.040
設備	
宿	.040
露天風呂	.039
最高	.030
夕食	.029
大変	.029

数値評価指標
風呂
部屋
食事
サービス
設備
立地

01_登別	02_草津	03_箱根	04_道後	05_湯布院					
良い	.058	温泉	.068	部屋	.078	温泉	.054	美味しい	.062
風呂	.057	湯畑	.064	良い	.060	良い	.051	宿	.061
温泉	.049	風呂	.062	風呂	.053	ホテル	.042	風呂	.059
美味しい	.044	良い	.061	美味しい	.049	美味しい	.042	露天風呂	.050
スタッフ	.037	美味しい	.042	露天風呂	.048	立地	.026	温泉	.043
設備	.031	宿	.041	温泉	.043	よい	.025	スタッフ	.038
宿	.030	ない	.036	スタッフ	.034	大変	.023	最高	.031
露天風呂	.030	最高	.031	宿	.033	浴場	.022	家族	.027
最高	.028	スタッフ	.030	夕食	.030	残念	.022	大変	.026
夕食	.027	露天風呂	.028	残念	.024	夕食	.021	よい	.025
大変	.025								

B_ビジネス	数値評価指標
部屋	.131
ホテル	.095
立地	.045
ない	.044
広い	.038
綺麗	.038
便利	.038
フロント	.037
駅	.036
快適	.034

数値評価指標
風呂
部屋
食事
サービス
設備
立地

06_札幌	07_名古屋	08_東京	09_大阪	10_福岡					
部屋	.064	ホテル	.063	部屋	.066	ホテル	.061	部屋	.060
ホテル	.061	部屋	.058	ホテル	.054	立地	.039	ホテル	.060
立地	.052	フロント	.035	快適	.034	駅	.036	立地	.041
ない	.033	綺麗	.032	駅	.034	綺麗	.033	便利	.034
広い	.033	駅	.030	立地	.034	ない	.031	広い	.032
綺麗	.031	便利	.029	ない	.033	便利	.031	駅	.030
便利	.031	立地	.029	フロント	.032	フロント	.030	綺麗	.029
フロント	.030	快適	.027	便利	.031	広い	.030	近い	.028
駅	.030	フロント	.024	近い	.027	近い	.028	フロント	.026
快適	.029	近い	.022	広い	.026	残念	.026	大変	.026

Tips: 「ツール」→「外部変数と見出し」→「リスト」→「変数リスト=カテゴリー」を選択→「▽特徴語」→「選択した値」→「関連語検索画面」→「フィルタ設定」→「品詞=名詞, 形容動詞, 未知語, タグ, 形容詞, 名詞B, 形容詞B, 名詞C」を選択→「▽特徴語」→「一覧(EXCEL形式)」で連続実行

# 演習1 — 特徴語の集計例

- ・「表記ゆれを吸収する」を参考に、「お部屋」「部屋」の表記ゆれをまとめて、その結果を「抽出語リスト」で確認する
- ・テキスト中の特徴語を集計し、エリアによって特徴語がどう異なるかを比較することで、注目する項目の違いを確認する（→P.18）
  - ・カテゴリー「レジャー」と「ビジネス」を比較する
  - ・カテゴリー「レジャー」(or「ビジネス」) の 5エリアを比較する

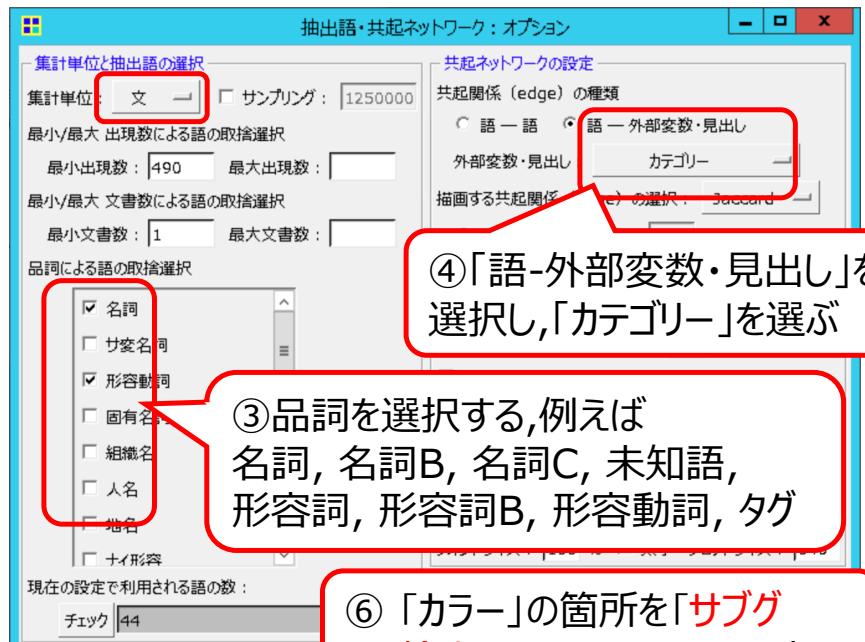
# Tips: 単語登録する

- 目的
  - 複数の単語に分かれる → 1単語として抽出できるようにする  
例) 「湯」「畠」の 2単語 →「湯畠」として 1単語
- 方法
  - 「前処理の実行」前に「強制出力する語の指定」に追加する
- 手順
  1. メニューから「前処理」「語の取捨選択」を選ぶ
    - 「強制出力する語の指定」欄に抽出したい単語を登録する
    - 「OK」ボタンで画面を閉じる
  2. メニューから「前処理」「前処理の実行」を選ぶ

# 使い方 – 共起ネットワークの作成3

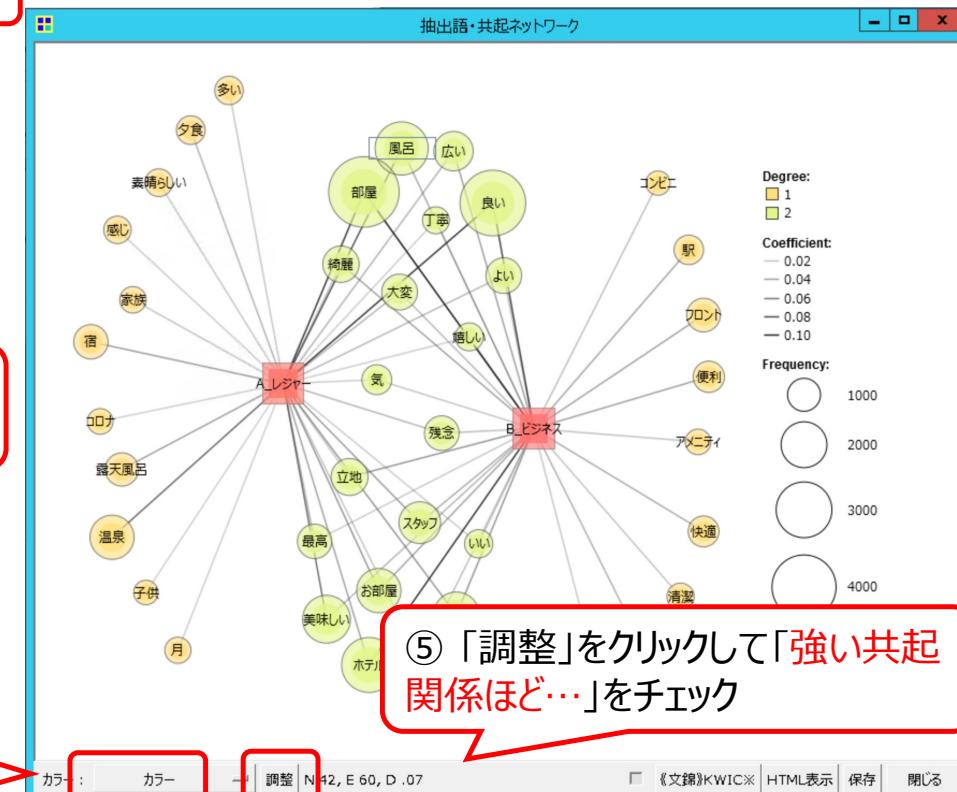
①メニューから「ツール」「抽出語」「共起ネットワーク」を選ぶ

②「集計単位」として「文」を選んで「OK」をクリック



③品詞を選択する,例えば  
名詞, 名詞B, 名詞C, 未知語,  
形容詞, 形容詞B, 形容動詞, タグ

⑥「カラー」の箇所を「サブグラフ検出(modularity)」に変更



# 実践編 — 特徴語の共起ネット

- 宿泊客は、どの項目のどこに注目しているか？
  1. カテゴリー「レジャー」と「ビジネス」を比較する
  2. カテゴリー「レジャー」(or「ビジネス」) の 5エリアを比較する

- 手順
  - 特徴語の共起ネットワーク図を作成

「ツール」→「抽出語」→「関連語検索」→「#直接入力[and]“<>エリア-->01\_登別”」「集計単位:文」  
→「フィルタ設定」→「品詞=名詞,未知語,タグ,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→  
「共起ネット」→「調整:**上位120,共起関係ほど濃い線に**」

- エリアによって特徴語(とその背景)がどう異なるかを比較
- 注目する項目の違いを考察する

# 直接入力: [and] の右側に入力する条件

レジャー:

<>カテゴリ-->A\_レジャー

<>エリア-->01\_登別

<>エリア-->02\_草津

<>エリア-->03\_箱根

<>エリア-->04\_道後

<>エリア-->05\_湯布院

ビジネス:

<>カテゴリ-->B\_ビジネス

<>エリア-->06\_札幌

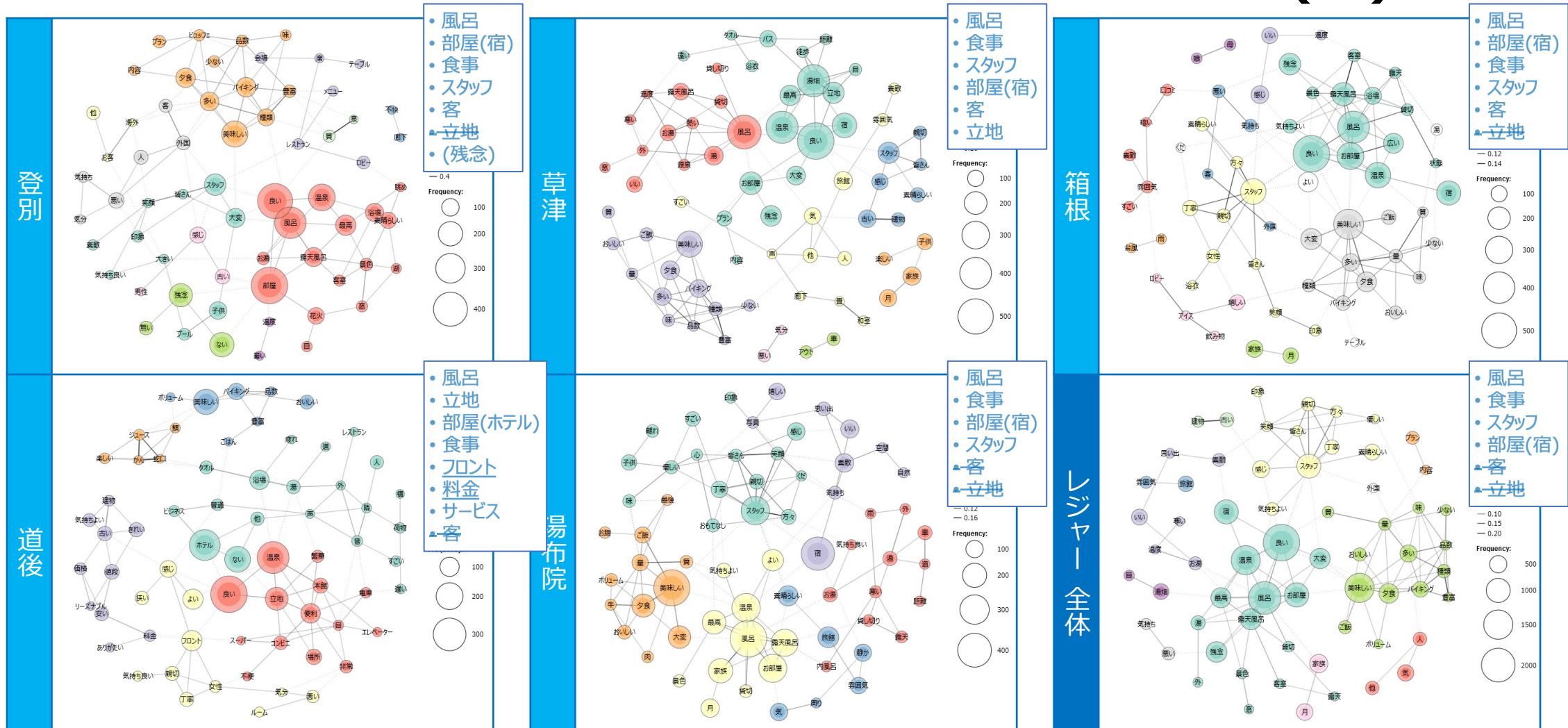
<>エリア-->07\_名古屋

<>エリア-->08\_東京

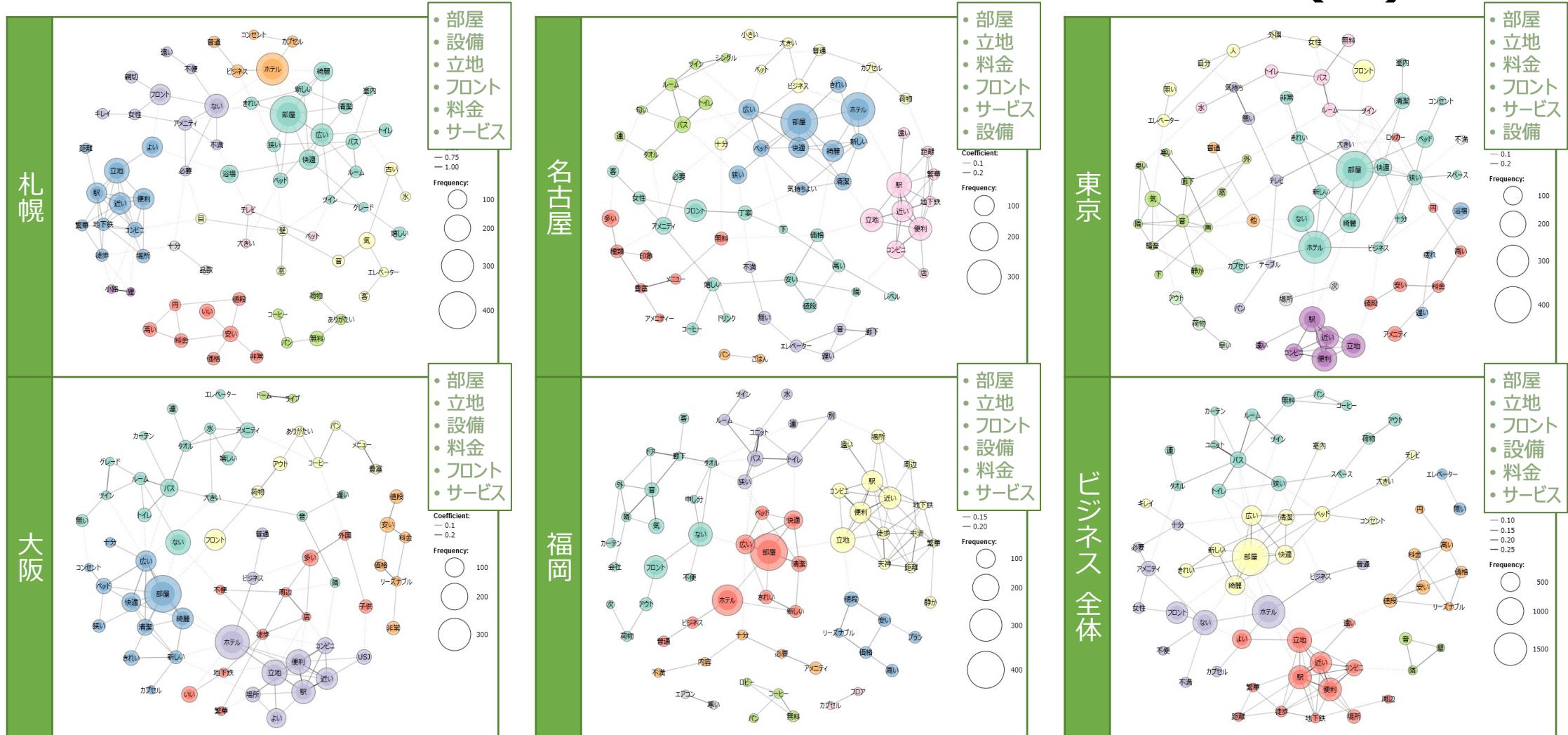
<>エリア-->09\_大阪

<>エリア-->10\_福岡

# 実践的な分析 — 共起ネットの出力例(1)



# 実践的な分析 — 共起ネットの出力例(2)



# 演習2 — 特徴語の共起ネット

- 特徴語の共起ネットワーク図を作成し,エリアによって特徴語(とその背景)がどう異なるかを比較することで,注目する項目の違いを確認する  
(→P.24,25)
  - カテゴリー「レジャー」と「ビジネス」を比較する
  - カテゴリー「レジャー」(or「ビジネス」) の 5エリアを比較する

# 結果の整理

- ・宿泊客が、どの項目のどこに注目しているかを列挙する

- ・エリアごとに、注目ポイントを列挙
- ・エリアごとで、注目ポイントを「好評」と「不評」に分類

カテゴリー	エリア	好評	不評
レジヤー	XXX	<ul style="list-style-type: none"><li>・風呂が広い</li><li>・...</li></ul>	<ul style="list-style-type: none"><li>・エアコンが臭い</li><li>・...</li></ul>

# 実践編 — 改善案を提案する(1/2)

- ユーザーは何をどう高評価しているか?
  1. カテゴリー「レジャー」と「ビジネス」を比較する
  2. 対照的な2エリアを比較する
- 手順
  - 特徴語とポジティブ意見の共起ネットワーク図を作成

「ツール」→「抽出語」→「関連語検索」→「#直接入力(and)"<>エリア-->01\_登別"」「Search Entry:\*ポジ」「AND検索」「集計単位:文」→「フィルタ設定」→「品詞=名詞,未知語,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:上位=120,共起関係ほど濃い線に」

- エリアによってポジティブ意見(とその背景)どう異なるかを比較
- 何がどう評価されているかを考察する

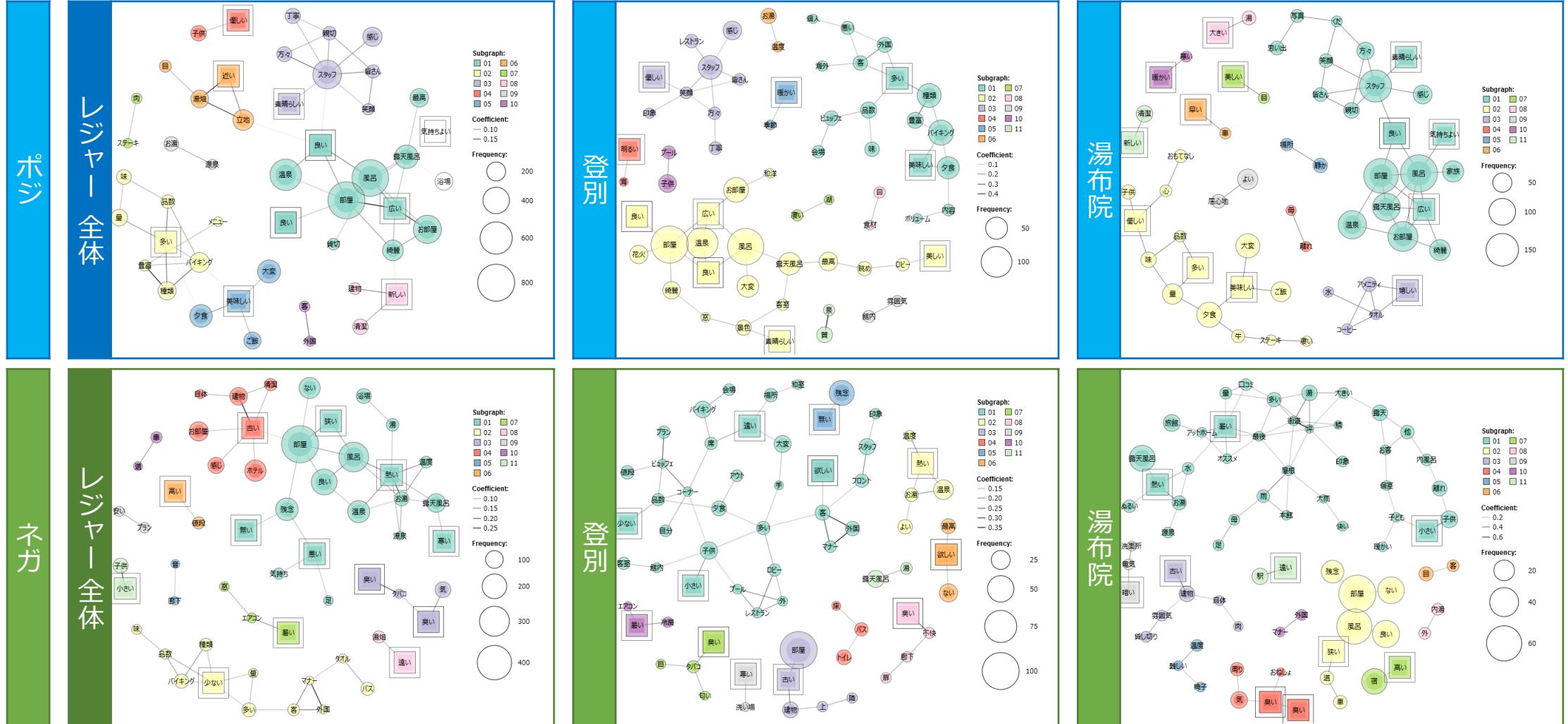
# 実践編 — 改善案を提案する(2/2)

- ユーザーは何をどう**低評価**しているか?
  1. カテゴリー「レジャー」と「ビジネス」を比較する
  2. 対照的な2エリアを比較する
- 手順
  - 特徴語と**ネガティブ意見**の共起ネットワーク図を作成

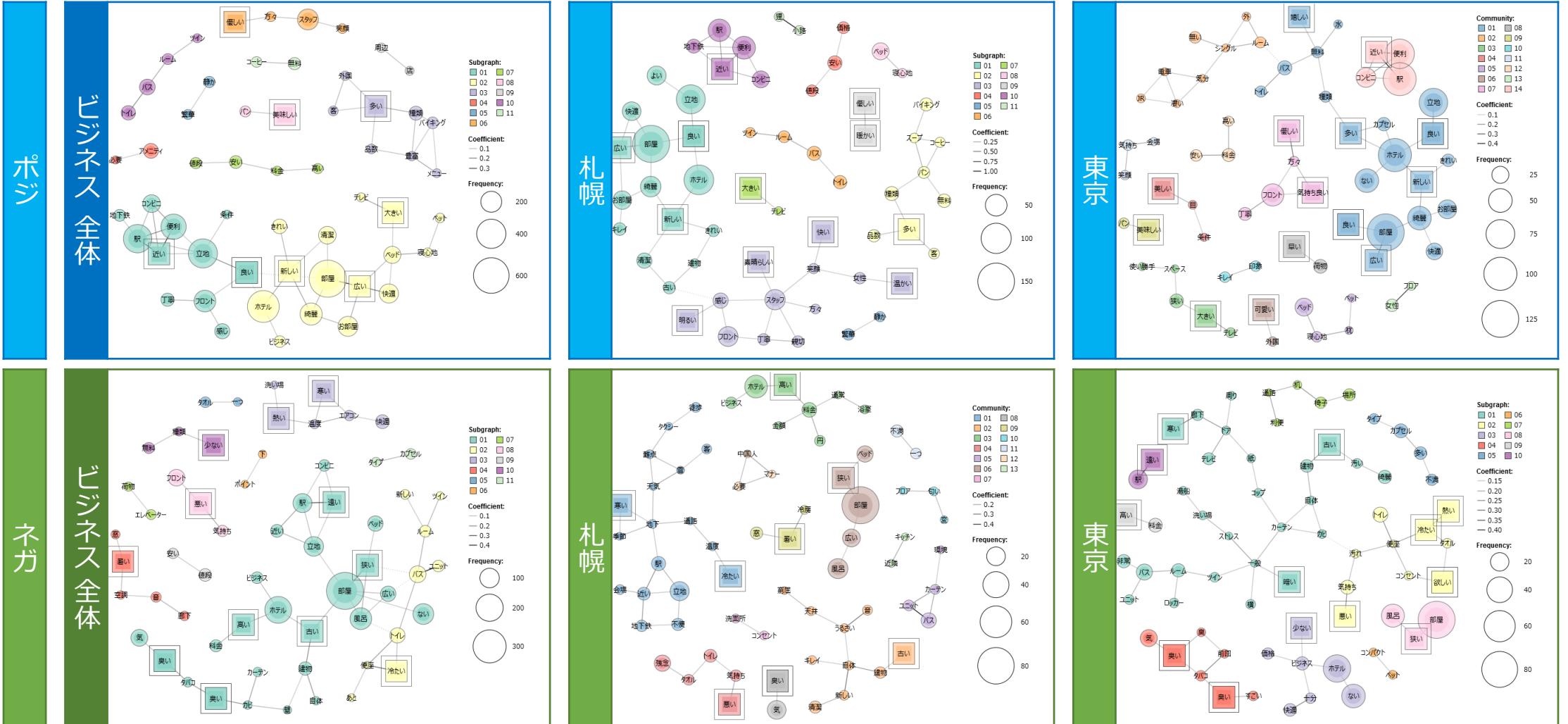
「ツール」→「抽出語」→「関連語検索」→「#直接入力(and)"<>エリア-->01\_登別"」「Search Entry:\*ポジ」「AND検索」「集計単位:文」→「フィルタ設定」→「品詞=名詞,未知語,形容詞,名詞B,形容詞B,名詞C」を選択→「集計」→「共起ネット」→「調整:**上位=120,共起関係ほど濃い線に**」

- エリアによって**ネガティブ意見**(とその背景)どう異なるかを比較
- エリアの課題を考察する

# 実践編 — 登別と湯布院のポジネガ比較



# 実践編 — 東京と札幌のポジネガ比較



# 演習3 — 改善案を提案する

- 特徴語とポジティブ意見の共起ネットワーク図を作成し,エリアによってポジティブ意見(とその背景)どう異なるかを比較することで,何がどう評価されているかを確認する (→P.24,25)
- 特徴語とネガティブ意見の共起ネットワーク図を作成し,エリアによってネガティブ意見(とその背景)どう異なるかを比較することで,何がどう評価されているかを確認する (→P.30,31)
  - カテゴリー「レジャー」と「ビジネス」を比較する
  - 対照的な2エリアを比較する

# 結果の整理

- ・主張を支持する図とユーザーの生の声(原文)を使って議論する
  - ・エリア X が評価されている点は何か?
  - ・エリア Y の課題は何か?
  - ・エリア Y の改善に向けた提案?

対象エリア	エリアX の評価ポイント	エリアY の課題	エリアYの改善案
エリアX: XXX エリアY: XXX	・風呂が広い 根拠原文: ... ・...	・エアコンが臭い 根拠原文: ... ・...	・... ・...

# 課題 – 実践的な分析

- 以下の 3点をスライドにまとめ **PDF ファイルで提出** してください
    - 演習1で作成した **特徴語のEXCEL集計**
    - 演習2で作成した **共起ネットワーク図** と **結果の整理**
    - 演習3で作成した **共起ネットワーク図** と **結果の整理**
- 形式: PDF, 提出先: manaba, 期限: 次週開始時刻(～18:20)

# Q&A

# 参考書

## (KH Coder)

- [1] 樋口耕一. 社会調査のための計量テキスト分析—内容分析の継承と発展を目指して【第2版】 KH Coder オフィシャルブック. ナカニシヤ出版, 2020.
- [2] 樋口耕一. テキスト型データの計量的分析 —2つのアプローチの峻別と統合一. 理論と方法, 数理社会学会, 2004, 19(1): 101-115.
- [3] 牛澤賢二. やってみよう テキストマイニング —自由回答アンケートの分析に挑戦!. 朝倉書店, 2019
- New** [4] 樋口耕一. 動かして学ぶ! はじめてのテキストマイニング: フリー・ソフトウェアを用いた自由記述の計量テキスト分析 KH Coder オフィシャルブック II.ナカニシヤ出版, 2022.

## (Windows環境によるデータ収集方法の参考に)

- [5] テキストマイニングソフトを利用した新未来洞察手法の研究. 第10分科会, (財)市場創造研究会. [http://www.shijo-sozo.org/news/第10分科会\\_1.pdf](http://www.shijo-sozo.org/news/第10分科会_1.pdf)

# 参考書

## (Rを使った参考書)

- [6] 金明哲. "テキストデータの統計科学入門." 岩波書店, 2009.
- [7] 石田基広. "RMeCabによるテキスト解析. Rによるテキストマイニング入門." 森北出版, 2008, 51-82.

## (他のツールを使った参考書)

- [8] 那須川哲哉. "テキストマイニングを使う技術/作る技術: 基礎技術と適用事例から導く本質と活用法." 東京電機大学出版局, 2006.
- [9] 上田隆穂, 黒岩祥太, 戸谷圭子. "テキストマイニングによるマーケティング調査." 講談社, 2005.

## (統計解析を中心とした参考書)

- [10] 前田忠彦; 山崎誠. 言語研究のための統計入門. くろしお出版株式会社, 東京, 2013.