

研究成果の社会還元

得られた新しい知見や分析技術を解説書にまとめ、札幌市立大学看護学部の学生等を対象に、試験的なテキストマイニング分析演習を実施し、解説書の精度向上をはかった。

今後は、本解説書を製本し、地域創生関係者を対象としたテキストマイニング講習会の実施を通して、地域創生関係者への分析技術の伝授を図る計画である。

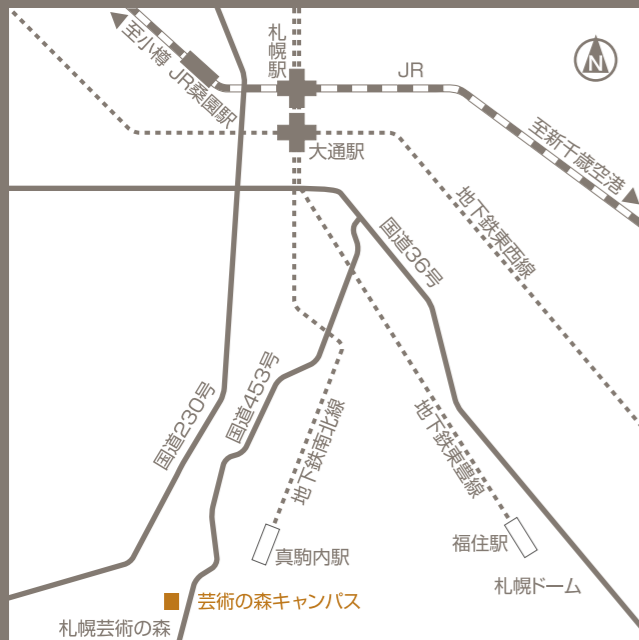


日本学術振興会科学研究費助成事業(学術研究助成基金助成金)(基盤研究(C)(一般)) 課題番号16K00713

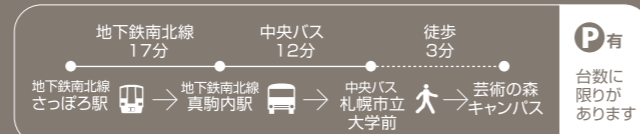
テキストマイニング分析手法による『地域創生デザイン』志向潜在知識抽出に関する研究

研究代表者：城間 祥之 (札幌市立大学 デザイン学部 教授)
共同研究者：柿山 浩一郎(札幌市立大学 デザイン学部 准教授)

質的データを量的データ化する



芸術の森キャンパス 〒005-0864 札幌市南区芸術の森1丁目



Report 2016-2018

研究目的

本研究の目的は、地域創生に関わる住民などの意識調査データ(テキスト)の分析にテキストマイニング手法を適用し、関心事や意見等の表出に係るキーワードなどを量的・客観的に視覚化することで、地域創生デザインに関する潜在意識などを抽出することである。

研究目的を達成するためには、膨大なテキストデータをふるいにかけて、特定のことばの表出頻度や、ことばとことばの係り受け関係などを導出すること、および導出されたことば関係データの意味を解釈することが必須となる。

研究成果

地域創生に係わる潜在知識抽出に効果的な分析プロセス

テキストマイニング用ソフトウェア:Text Mining Studioを例に

- 回答文書データファイル(csv file)をText Mining Studioへ読み込み「分かち書き」を実行する。**
文章は形態素(言語で意味を持つ最小単位)に区切られ、形態素ごとの品詞が確定された分析のためのオリジナルテキストに変換される。
- 「テキスト情報ツール」を用いて分析する。**
基本情報(総行数、平均行長(文字数)、総文章数、平均文章長(文字数)、延べ単語数、単語種別数)や品詞出現回数などが算出される。
- 「ビジュアル集計ツール」を用いて分析する。**
属性ごとの回答者数が算出される。
- 「単語頻度解析ツール」を用いて分析する。**
単語の出現頻度が算出される。
- 「係り受け頻度解析ツール」を用いて分析する。**
主語と述語のような係り受け表現から、回答者が記述したポジティブ(またはネガティブ)な表現事例を抽出する。
- 「評判抽出ツール」を用いて分析する。**
ポジティブ(またはネガティブ)な表現に用いられた単語(名詞や自立動詞)の頻度が算出される。
- 「ことばネットワーク分析ツール」を用いて分析する。**
矢印付き有向線で構成される係り受け関係ネットワーク図が出力される。
- 「対応パブル分析ツール」を用いて分析する。**
ことばを介した属性同士の関係図が出力される。

具体事例

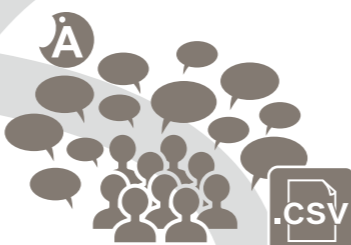


「札幌市南区在住65歳以上高齢者の健康に関するニーズ調査」

札幌市南区在住65歳以上高齢者、約9000人を対象に実施健康ニーズ調査への回答者数は2998人(33.3%)
問21への有効回答者数は892人(回答者の29.8%)

問21

「旧真駒内緑小学校」の跡地を利用し、札幌市と札幌市立大学が協働してCOCキャンパス(同封資料参照)を作る予定です。南区の住民として、そこで行って欲しいことや期待することを何でもお書きください。



テキスト列の基本情報の算出

項目	値
総行数	892
平均行長(文字数)	57.5
総文章数	1641
平均文章長(文字数)	31.3
延べ単語数	10887
単語種別数	4296

一つの文章の文字数を表してあり、簡潔な文章(回答)だと判断可能。

テキスト情報ツール
基本情報を得る

品詞	回数(%)
名詞	7512 (69.00%)
代名詞	169 (1.55%)
動詞	1610 (14.7%)
形容詞	499 (4.58%)
形容動詞	333 (3.06%)
連体詞	105 (0.96%)
副詞	408 (3.75%)

感性を表す単語は少なく感情的な回答は少なかったと判断可能。

ビジュアル集計ツール
属性情報を得る

テキスト列以外の属性情報の算出

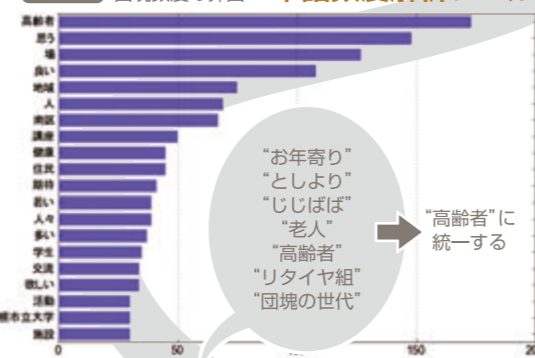
年齢	男性	女性	未記入	合計
1:65歳~69歳	178	173	1	352
2:70歳~74歳	138	84	0	222
3:75歳~79歳	100	53	0	153
4:80歳~84歳	65	40	0	105
5:85歳以上	34	25	0	59
年齢未記入	1	0	0	1
合計	516	375	1	892

年齢が上がるにしたがって男女ともに回答者数が漸減していることが読み取れる。



単語の出現頻度の算出

出現しているキーワードを得る
単語頻度解析ツール



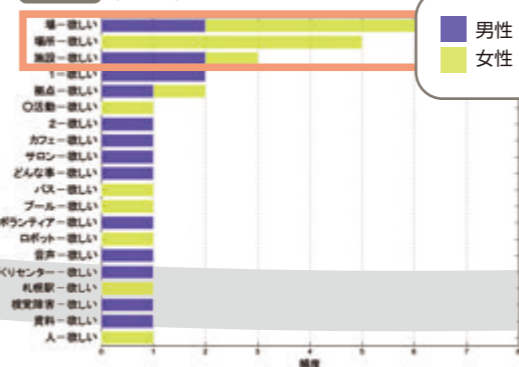
ことばをまとめあげる類義語辞書

気になることばや態度、要望を得る
係り受け頻度解析ツール

問21「...COCキャンパス...で行って欲しいこと...」と問うているので、回答文における気になることばを、「欲しい」に絞込む。



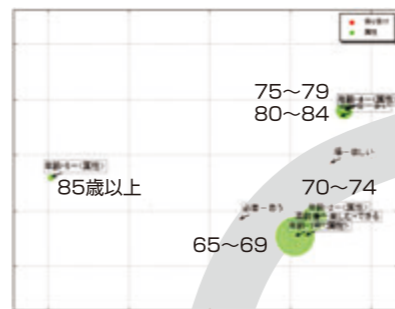
「欲しい」の係り受け表現(男女別)の算出



高齢者が活動する「場」や「場所」に関する意見(要望)が多く女性の方が地域での交流や活動の場を積極的に求めていると判断可能。



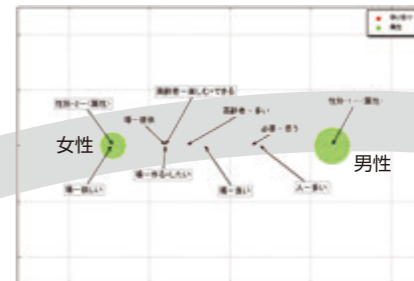
ことばを介した特徴の可視化[年齢区分]



「南区一発展」という係り受け表現は、どの係り受け表現とも距離が大きく離れている。これらの位置関係から区分5(85歳以上)の意見(要望)は他の年齢区分のいずれの意見(要望)ともかなり離れていると判断可能。



ことばを介した特徴の可視化[性差]



男性の丸印は女性の丸印と比較して大きく、男性の意見(要望)が女性よりも多いと判断可能。男性は人・モノ・ことの多少や必要と関連する表現が多いことが読み取れる。女性はどのような活動をしたいのか、どのような場がほしいのかが明確であると判断可能。



大量の文書データに対して...



ふるいにかける感度調整技術力



金塊を見つけ出す能力

(重要キーワードや関係性)



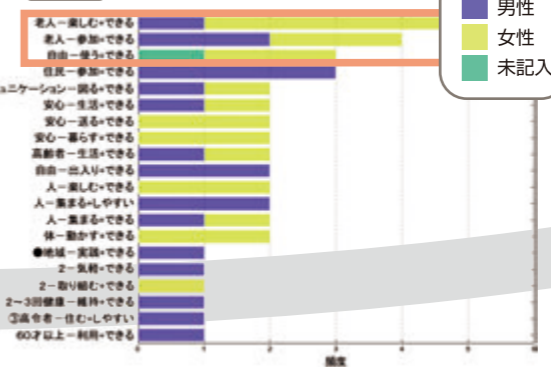
金塊から潜在知識を抽出する能力

「ない(否定的)」の係り受け表現(男女別)

お金にかかる+ない」は、お金がかからないことを表現した意見(要望)であり、否定的な係り受け表現に分類されているが、肯定と判断する。



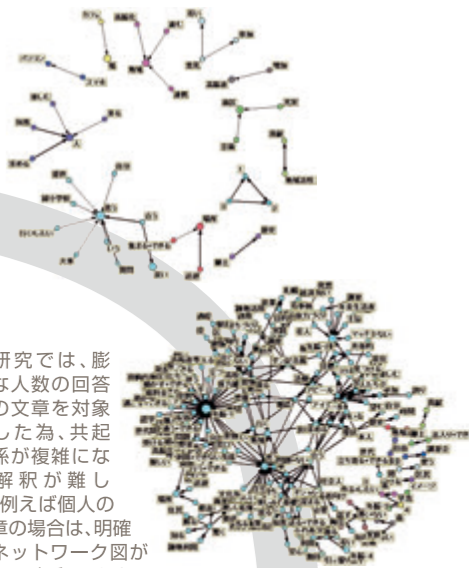
「できる(肯定的)」の係り受け表現(男女別)の算出



【女性の傾向】楽しく参加できる活動や自由に使うことができる場・場所についての肯定的な意見(具体的要望)が多い
【男性の傾向】住民一参加+できる「など、概念的な意見が多い」と判断可能。



共起関係にある(一文中に同時に出現する)単語同士の関係の可視化

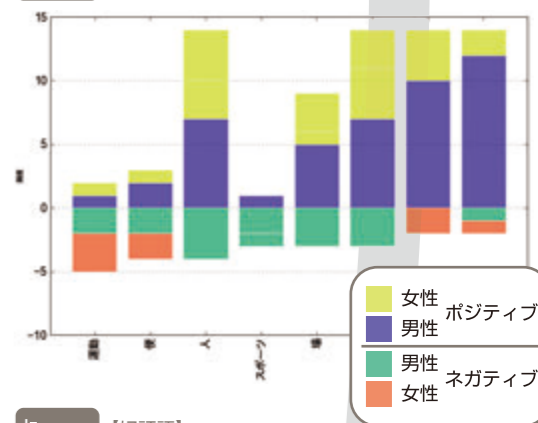


*本研究では、膨大な人数の回答者の文章を対象とした為、共起関係が複雑になり解釈が難しい。例えば個人の文章の場合は、明確なネットワーク図が得られ、解釈しやすい。

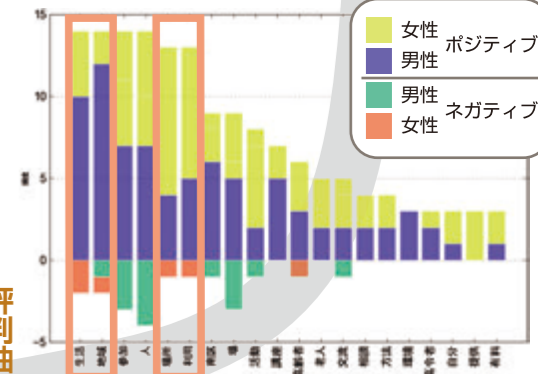
話題単位でデータ全体を概観することばネットワーク分析ツール



【不評語】抽出されたことば(単語)に関する否定的な意見の算出



【好評語】抽出されたことば(単語)に関する肯定的な意見の算出



「地域」や「生活」に関する肯定的な意見(要望)は男性が圧倒的に多い。「場所」、「利用」、「活動」に関する肯定的な意見(要望)は女性に多い。よって、男性は社交的な場や活動に消極的であると判断可能。

評判抽出ツール
ポジティブ/ネガティブな意見を得る