

# 大容量文書データのテキストマイニング分析

～札幌市南区在住65歳以上高齢者の意識調査データを対象として～

札幌市立大学 デザイン学部  
城間祥之・柿山浩一郎

## 目的

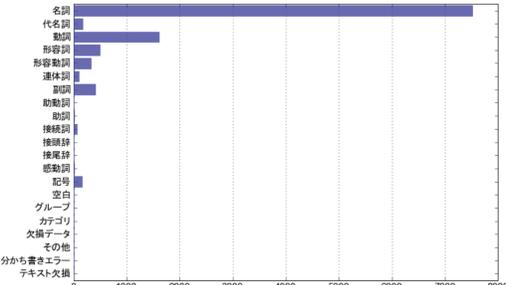
本研究の目的は、地域創生に関わる住民などの意識調査データ(テキスト)の分析にテキストマイニング手法を適用し、関心事や意見等の表出に関わるキーワードなどを量的・客観的に視覚化することで、地域創生に関する潜在意識などを抽出することである。

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、および導出された関係データの意味を解釈することが必須となる。

ここでは、文部科学省「平成25年度「地(知)の拠点整備事業」に採択された札幌市立大学『ウェルネス x 協奏型地域社会の担い手育成「学び舎」事業』の一環として実施された「札幌市南区にお住いの65歳以上の方の健康に関するニーズ調査」の集計データから、「札幌市立大学COCキャンパスで行って欲しいことや期待することをお書きください」との質問に対する回答を抜き出し、分析したので報告する。

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

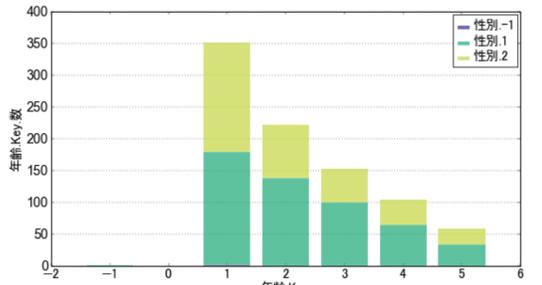
基本情報



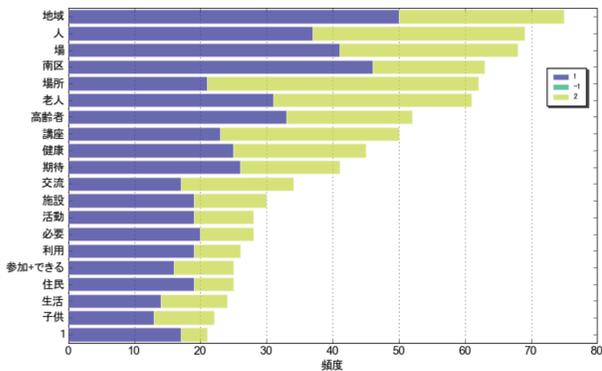
品詞別出現回数

年齢	男性	女性	性別未記入	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

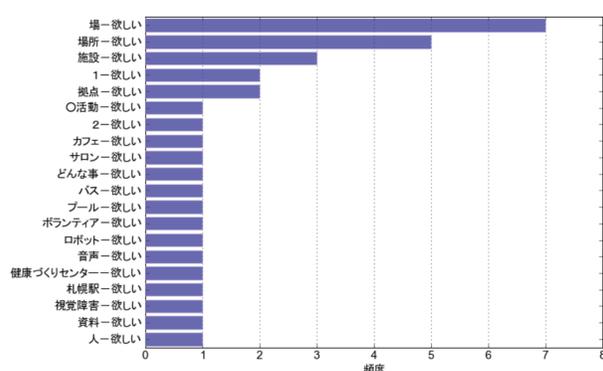
回答者の年齢 x 性別クロス集計と欠損データの関係



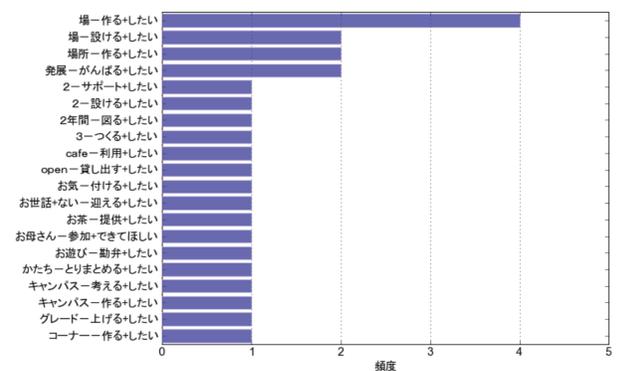
年齢区分別回答者数



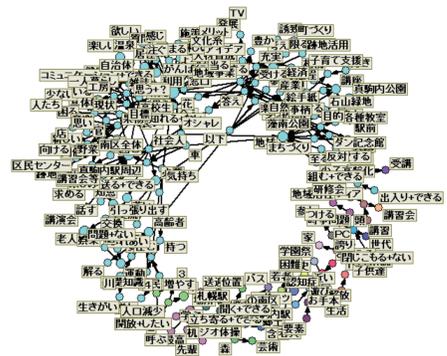
単語頻度解析(品詞:名詞・形容詞・形容動詞・動詞)



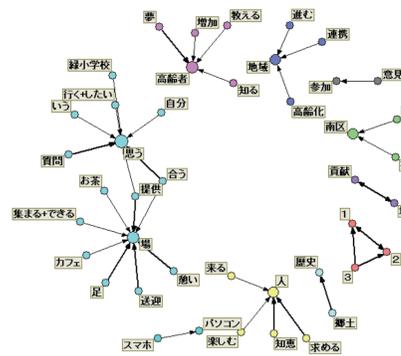
係り受け頻度解析(単語フィルタ:係り先単語:”欲しい”を含む)



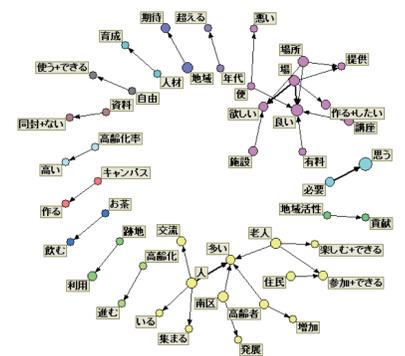
係り受け頻度解析(態度表現フィルタ:係り先態度表現:要望)



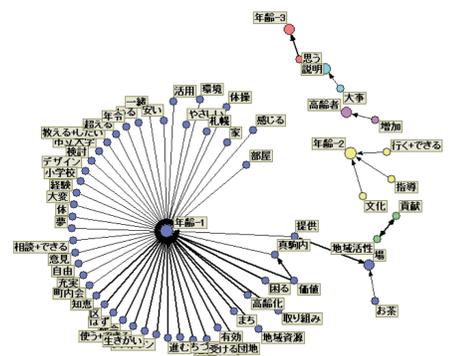
ことばネットワーク:品詞設定:話題一般  
(行単位での共起、出現回数:2回以上)



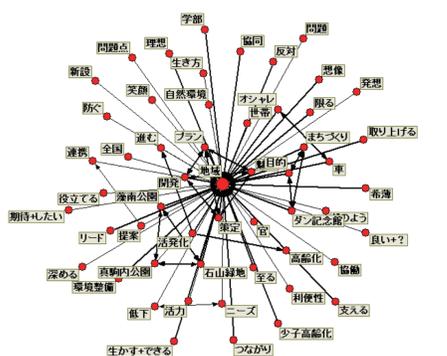
ことばネットワーク:品詞設定:行動  
(行単位での共起、出現回数:4回以上)



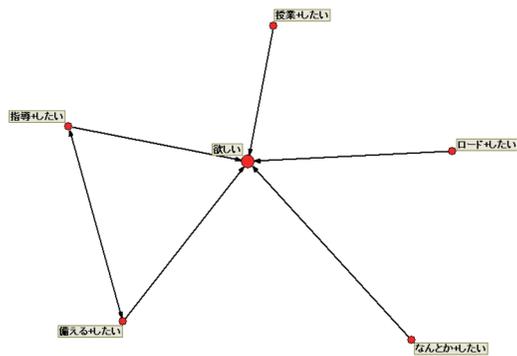
ことばネットワーク:品詞設定:話題一般  
(係り受け頻度回数:3回以上)



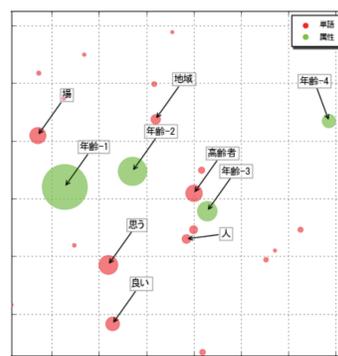
ことばネットワーク:品詞設定:話題一般  
文書単位での共起、出現回数:5回以上、属性設定:年齢



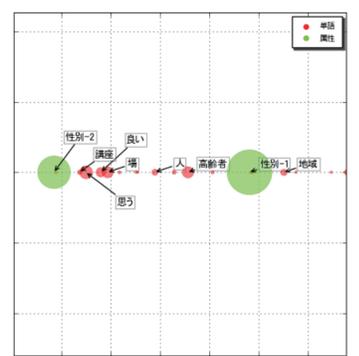
注目語”地域”と一緒に出現することば



注目語”欲しい”と一緒に出現することば



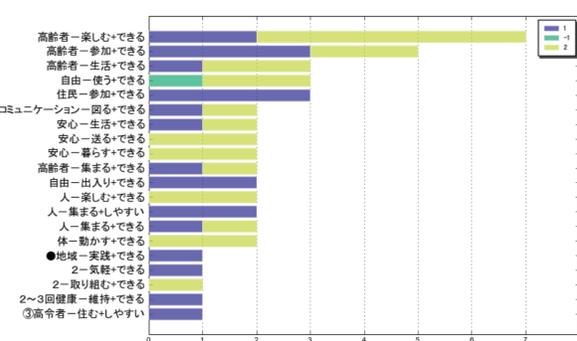
対応バブルチャート①  
(年齢区分とことばの関係)



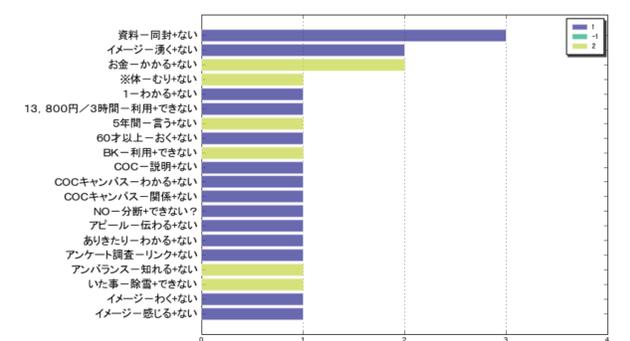
対応バブルチャート②  
(性別とことばの関係)



評判抽出(好評語ランキング)



係り受け頻度解析(ポジティブな意見)



係り受け頻度解析(ネガティブな意見)