

January 31, 2020

# 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

1615052 山元 悠貴

富山県立大学 情報基盤工学講座

1. はじめに
2. 発想支援システムとは
3. データドリブンによる最適化
4. おわりに

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

# 1. はじめに

2/10

## 背景

人間の創造活動において精神的な労力が必要である。この精神的な労力の負担を減らすために発想支援についての研究がされている。現在、情報処理技術の発達に伴って発想支援についての研究も進んでいる。これからの時代はよりアイデア発想が重要になると考える。既存の発想支援には一つのキーワードに対して共起ネットワークを作成するシステムなどがある。

## 目的

- 1 複数のキーワードを入力とした発想支援
- 2 キーワードに対して関連語を 3D 有向グラフにして表示するシステムの作成

2/10

## 2. 発想支援システムとは

3/10

### 発想のプロセス

- 1 発散的思考
- 2 収束的思考
- 3 アイデア結晶化

本研究の発想支援システムではこのプロセスに従ってシステムを構築する。

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

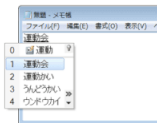
## 2. 自然言語処理の例

4/10

### 身近な例

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

#### ■文字変換



#### ■検索エンジン



#### ■機械翻訳



#### ■対話システム



## 3. テキストマイニング

5/10

### テキストマイニング

テキストマイニングとは、大量かつ多量なデータを様々な観点から分析し、役に立つ情報を取り出そうとする技術である。

### サイバー空間からのテキストデータ収集

発想支援においてキーワードからより関連度の高い単語を事を考える。

そこで、より良いデータを多く収集するためにインターネット上のテキストデータを収集することとする。

### 自然言語処理

自然言語処理とは私たちが使う言語をコンピュータに処理させる技術のこと。

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

## 3. PageRank

6/10

### PageRank

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|} \quad (1)$$

$r(P_j)$ : page  $P_j$  のページランク

$P$ : 全ページの集合

$B_P$ : 入リンクの集合

$|P_j|$ : ページ  $P_j$  の出リンク数の総数

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

## 4. アルゴリズム

7/10

### アルゴリズム

- 1. はじめに
- 2. 発想支援システムとは
- 3. テキストマイニング

- ① 複数のキーワードから Google の検索結果の X ページ分の URL を保存
- ② URL からそのページの HTML を取得して HTML タグを消去したテキストデータを取得
- ③ そのテキストデータに対して自然言語処理の前処理を行った後、形態素解析し、品詞を絞ってリストにする
- ④ そのリストから単語の順序に応じた隣接行列を作る
- ⑤ 隣接行列の規模が大きすぎるので PageRank を使って関連性の高い単語に絞る
- ⑥ 隣接行列を 3D 有向グラフに適応する

## 隣接行列を 3D 有向グラフに使える形にしたデータ

1. はじめに
2. 発想支援システムとは
3. テキストマイニング

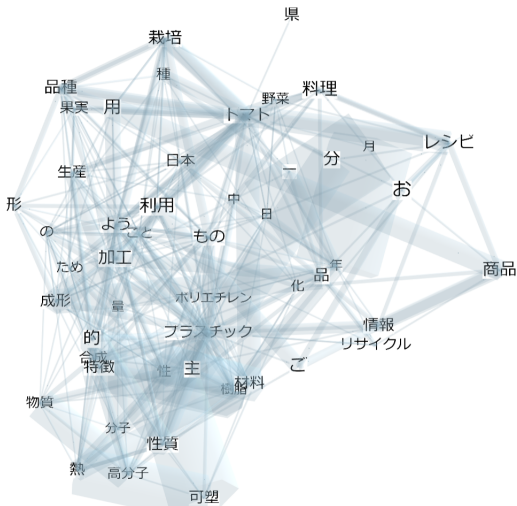
[illegible]



## 4. 出力結果

9/10

トマトとプラスチックを入力したときの出力結果



1. はじめに
2. 発想支援システムとは
3. テキストマイニング

## まとめ

- ① 複数のキーワードに対する関連語を **3D** 有向グラフにして発想支援システムとして提案した

## 今後の課題

- ① 単語のベクトル化などの自然言語処理を適応する
- ② キーワード間の最長経路を求めることでその間の関係性が可視化できる可能性がある
- ③ 単語の順序関係を係り受け解析を使って隣接行列にするときの精度を良くする
- ④ 処理時間が **3** 分くらいかかるので短くする