

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 (仮)

富山県立大学 工学部 電子・情報工学科
1715088 渡辺健太

担当教員：奥原浩之

0 キーワード

テキストマイニング, 共起ネットワーク, 自然言語処理, 発想支援

1 はじめに

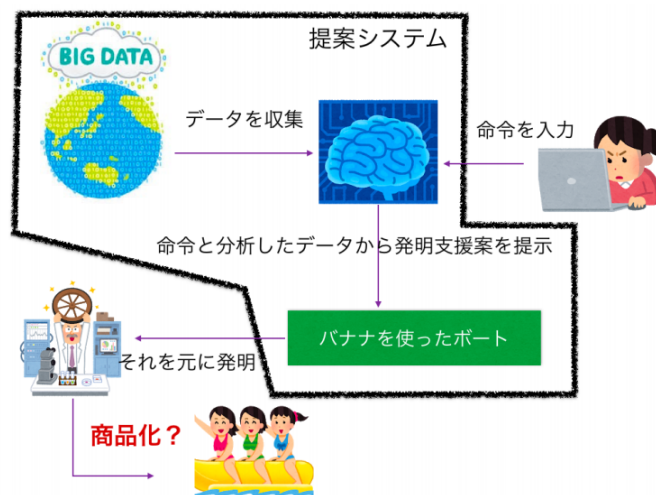
人間の創造活動において精神的な労力が必要である。この精神的な労力の負担を減らすために発想支援についての研究がされている。現在、情報処理技術の発達に伴って発想支援についての研究も進んでいる。これからの時代はよりアイデア発想が重要になってくると考える。既存の発想支援には一つのキーワードに対してそのキーワードに関連するテキストデータから共起ネットワークにするものがある。

2 基本知識

本研究内で前提となる基本的な項目についての説明。

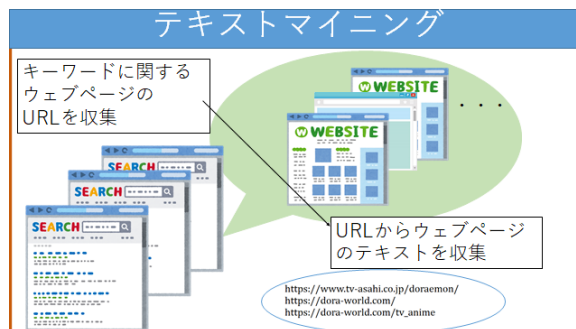
2.1 発想支援

現代社会においてインターネット上の情報量は莫大になっており、今後も増え続けることが予想される。このインターネット上の情報を収集して分析することで発想支援に生かせると思う。発想支援において重要なことはキーワードからより関連度の高い単語をより多く表示させることである。そこで、より良いデータを多く収集するためにサイバー空間からテキストデータを収集することとする。



2.2 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。インターネット上のテキストを用いることで大量のデータを活用することができる。まず、収集した文章に対して HTML タグや JavaScript のコードを取り除くクリーニング処理をする。その次に形態素解析を行う。形態素解析とは文を形態素ごとに分解する技術である。発想支援において必要な名詞や動詞に分解することである。また、助詞の「は」や「が」助動詞の「です」は不要なので取り除く。そして単語群に対して正規化を行う。正規化することで文字種を統一できる。半角と全角や数字を統一することで同じ単語として扱うことができる。



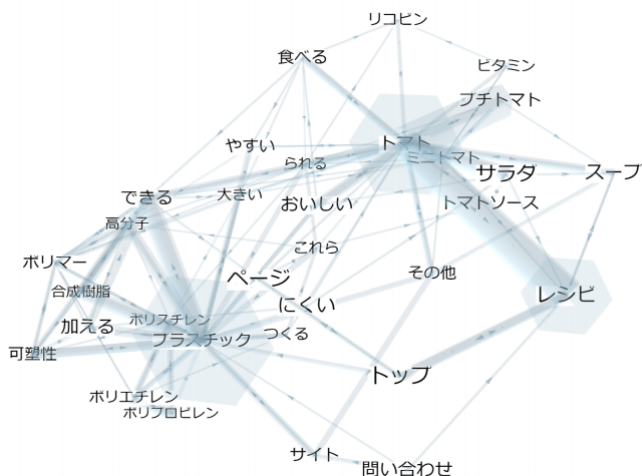
3 現在のシステムについて

3.1 アルゴリズム

- [1] 2つのキーワードから Google 検索の 1 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] 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] リストから単語の順序に応じた隣接行列を作る
- [5] 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

3.2 実行手順

- [1] サーバ (Server.py) を起動する
- [2] system.html を実行するとブラウザが開く
- [3] 任意のキーワードを 2 つ入力し、実行をクリック
- [4] 3D グラフが描画できれば成功、マウス操作で視点移動が可能



4 課題

- 発想支援への活用 (向き、ルート、…)
- システムの有効的な評価方法
- 様々な品詞での実装
- UI/UX の改善