

IP ランドスケープ支援のための 特許情報のベクトル化を用いた 共起ネットワーク作成システム

平井 遥斗

富山県立大学 情報システム工学科

2023 年 01 月 26 日

1.1 研究の背景

2/13

背景

近年，コロナウィルスの影響やグローバル化，インターネット技術やAI，IoT等のデジタル技術の進展，顧客のニーズの多様化などや社会環境などの急速な変化により，経営環境は大きく変化している。そういった VUCA な時代に，持続的な発展を図るためには，自社の核となる独自の強みを生かし，他者との差別化を図ることが極めて重要である。

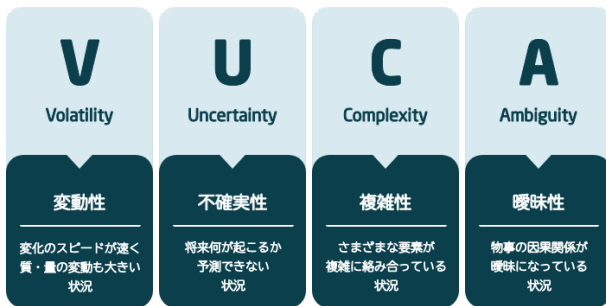


図 1: VUCA

1.2 研究の目的

3/13

目的

本研究の目的は、今日に至るまでに蓄積された莫大な量の特許情報を用いて過去から現在に至るまでの技術開発を可視化することで、特許探索を効率的に行えるようにすることで IP ランドスケープの実施を支援することである。

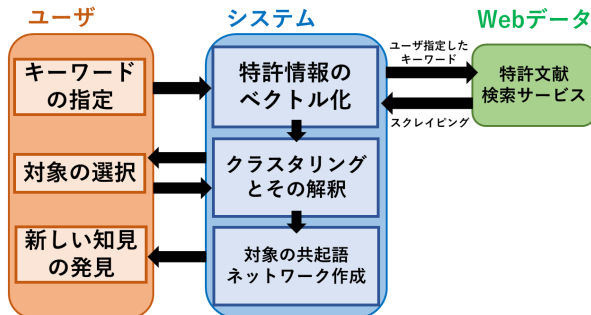


図 2: システムの流れ

1.2 研究の目的

4/13

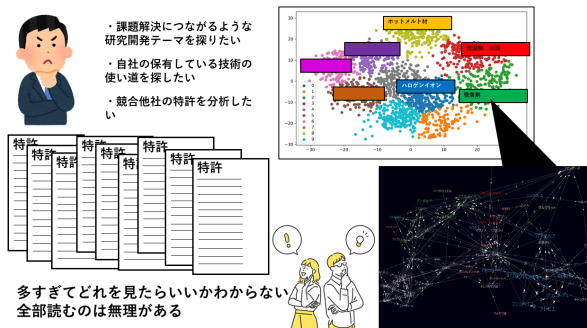


図 3: 目的

2.1 知的財産戦略

知的財産戦略とは企業が保有する知的財産を経営戦略の一環として取り入れ、企業の競争力を高め、事業目標を達成することを目的とする戦略である。知的財産戦略は、経営戦略と密接に関係しており、企業全体の戦略において各部門や機能の方向性を決定する重要な役割を果たしている。日本において、知的財産戦略は特許などの知的財産 (Intellectual Property:IP) と景観や風景を意味する (Landscape) を組み合わせた造語で「IP ランドスケープ」と呼ばれることが多い。

年数の指定

スクレイピングに時間がかかることと、入力するキーワードによって取得できる特許の数が違うことを解決するために、ユーザーが取得したい年数を指定できるようにした。指定できる単位は6年単位で6年、12年、18年、24年を選択できるようにした。



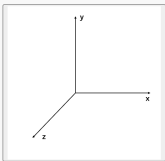
図 4: フロントページ

ネットワークの大きさ

前回、指摘があったので、共起ネットワークの大きさをユーザーが指定できるようにした。3D グラフ 2D グラフどちらも、大、中、小、の3つの大きさを指定できるようにした。

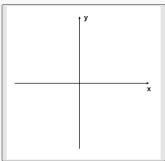
奥原研究室-Rene研... 新しいタブ 自分のファイル - One... Katamaril R FILMAGA いい映画と... bert 成績 成績 (D)

インバランス料金/発電側インバランス料金/需要側インバランス料金



3Dグラフ

☐ 小 ☒ 中 ☐ 大



2Dグラフ

☐ 小 ☒ 中 ☐ 大

特許一覧

[JP6841231B2](#)

[JP6845091B2](#)

[JP2019046155A](#)

[JP7037353B2](#)

ID771101607093A

クラスに含まれる特許の表示

クラスを選択したときに、それらの中に含まれる特許の一覧を表示できるようにした。リンクをクリックすることで、その特許に対応した GooglePatents のページに飛ぶことができ、そこから特許を見ることができる。

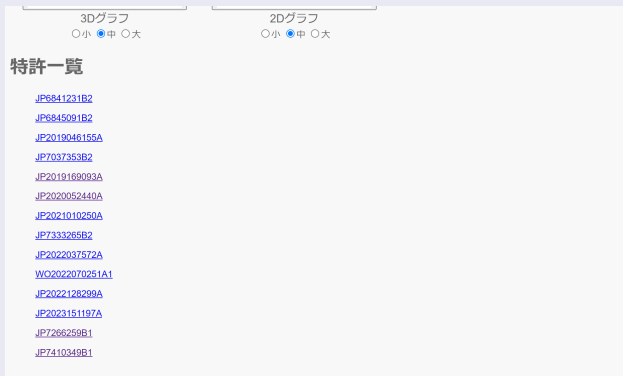


図 6: 特許一覧

Simpson 係数

共起関係の導出方法について

今まで導出方法として Jaccard 係数を用いて共起関係を導出していたが、Dice 係数、Simpson 係数を用いた導出も行って見た。

Jaccard 係数

2つの集合に含まれている要素のうち共通要素が占める割合を表しており、係数は0から1の間の値となる。

Dice 係数

Jaccard 係数では、片方の集合だけ要素数が膨大である場合に類似度が著しく下がる問題があった。それを解決するために、共通要素数を重視した類似度計算を行っている

Simpson 係数

Dice 係数よりもさらに要素数の影響を抑えたものが Simpson 係数。相対的に、共通要素数を重視した類似度計算を行っている。

しきい値

Simpson 係数の問題点

Simpson 係数では片方の集合の要素数が少ないときに、値が極端に 1 に近づいてしまう。

片方の要素が片方の真部分集合であるとき類似度が 1 となり、それらを正しく計算できているかは適用する問題によって異なる。

今回の場合ではお互いの要素数の差が 5000 以上のものについては除外している。

>5000 よりも小さくしてしまうと、出力される共起関係が少なくなってしまう。

Simpson 係数が 1 になる共起関係を除外する。

> 今回の場合共起関係が 1 になるということは起こりにくいと考えられる。1 になると考えられるケースはお互いの単語がもともとは 1 つの複合語であり、それらがうまく抽出できていない場合という風にとらえることができる。

数値実験

行う数値実験の内容について
システムを利用してもらい

- システムの使いやすさ
- ユーザーインターフェースの見やすさ
- グラフの見やすさ
- 操作性の良さ

などを評価してもらう。

特許のベクトル化とクラスタリングが正しく行われているかを検証する。

- 近い特許同士の内容が同じような内容になっているか。
- 遠い特許同士の内容が違う内容になっているか。

を検証する。

数値実験

並列処理による実行時間の変化の検証.

＞並列にする数を変化させたときのスクレイピングの時間の違いを検証する.

Simpson 係数のしきい値の変化によって作成される共起語ネットワークがどのように変わるのか.

＞ Simpson 係数のしきい値とお互いの集合の要素数の差の許容範囲を変えることで作成される共起語ネットワークにどのような違いが表れるのか.

クラスターのタイトルが正しく抽出できているか. またランダムで抽出したときとどのような違いが表れるか.

＞クラスターにつけられたタイトルが正しい解釈として利用できるのかを検証する. またランダムで行った時とユークリッド距離を用いた時とのタイトルの違いを検証する.

共起語ネットワークから実際にどのような知見を発見できるか.

＞作成された共起語ネットワークから新しい知見を発見する.

まとめ

- システムの細かいところの変更を行った.
- 本論を完成に近づけたい.
- 発表のスライドを完成させる.
- 数値実験を行う.