

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

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

Co-occurrence Word Network Creation System
Using Vectorization Patent Information
for IP Landscape Support

平井 遥斗 (Haruto Hirai)
u020032@st.pu-toyama.ac.jp

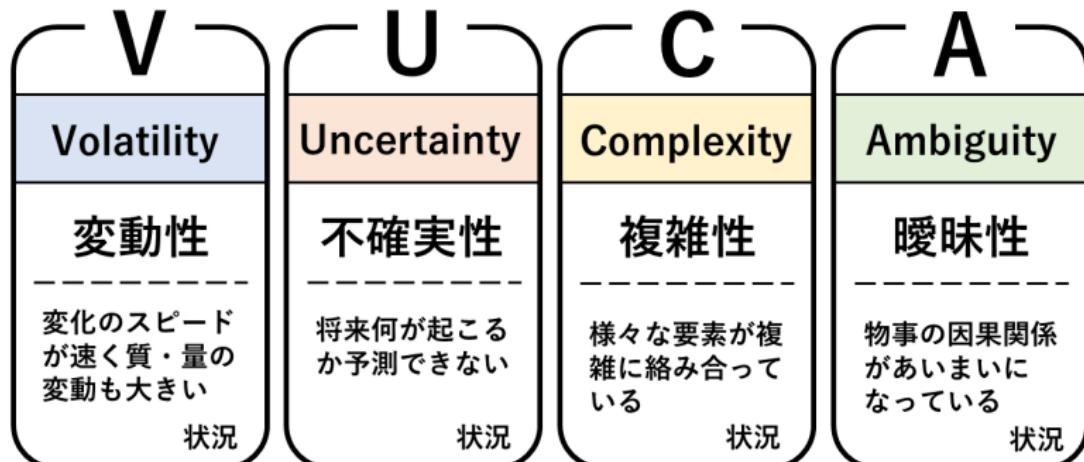
富山県立大学情報システム工学科 4 年

1.1 研究の背景

2/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

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、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。



1.2 研究の目的

- はじめに
- 知的財産戦略と特許情報処理
- 特許情報の可視化
- 提案手法
- 実験結果並びに考察
- おわりに

そこで、IP ランドスケープの支援を行うため、莫大な量の特許を効率的に探索を行い、その中から共起語ネットワークを用いて新たな知見を発見することができるシステムを作成する。

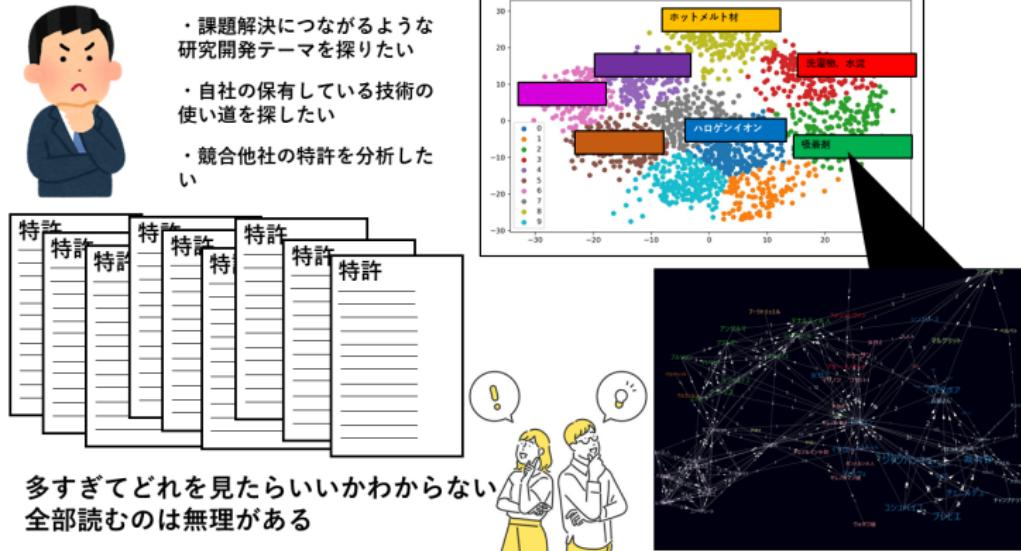


図 1: 目的

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

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。

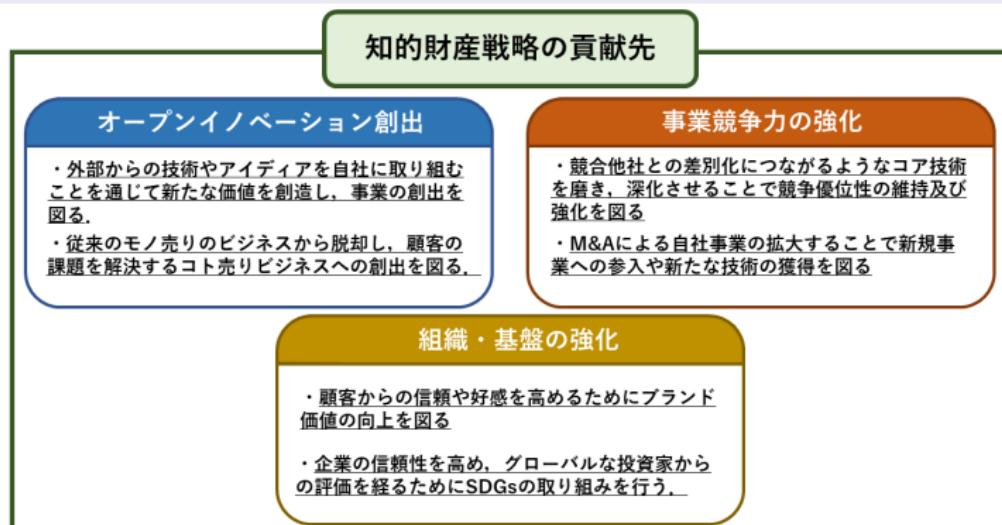


図 2: 活用先

2.2 特許情報処理と活用

5/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

特許情報処理とは、特許文章や特許データベースの情報をコンピュータで自動的に解析・処理を行うことで、莫大な量の特許情報から有用な知見を効率的に見出し、知的財産戦略の立案や技術動向の分析などに活用することを目的としている。

特許情報

研究開発の重複防止、既存技術の活用、無用な紛争の回避などに役立つ。

特許番号

特許として認められた発明に付与される7桁の番号
原則 審査段階、年、通し番号

で構成されている

特許プラットフォーム

特許情報データベースを提供するオンラインサービス。
各国の特許庁が運営するデータベース

Google Patents

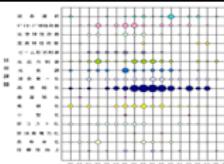
メーカーが提供するサービス

などがある。

特許番号などの基本情報での検索が可能

特許情報処理

特許文章やデータベースなどをコンピュータで自動的に解析・処理を行う



バブルチャート*

課題を時系列に並べ
その件数を

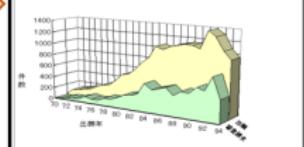
バブルの大きさ

により表示したもの

件数推移マップ**

技術開発の

- **開始時期**
- **急増期**
- **衰退期**



を捉えることができる

* http://www.jpo.go.jp/shiryou/s_sonota/map/denki06/map4.htm 特許庁ホームページより

** http://www.jpo.go.jp/shiryou/s_sonota/map/denki15/s/s-1-2.htm 特許庁ホームページより

図 3: 特許情報処理

2.3 テキストマイニングと自然言語処理

6/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

テキストマイニングを行うことで、顧客のニーズや市場動向を把握したり、新製品の開発やマーケティングの戦略を策定したりすることができる。テキストマイニングでは、自然言語処理の技術を活用して、テキストから意味のある情報を抽出しそれらの分析を行う。

テキストマイニング

定型化されていない文章から情報を抽出する技術
近年ではテキストデータの増加に伴い、様々なツールが提供されている

Selenium

- ・自動的に実際のブラウザを操作・制御することが可能。
- ・様々なサイトでデータを取得することができる。
- ・ブラウザの種類やバージョンに依存するため、互換性の問題が発生する可能性がある。

BeautifulSoup

- ・HTMLのデータ抽出に特化している。
- ・処理速度がSeleniumよりも高速である。
- ・ページの種類によっては取得できないこともある。

自然言語処理

コンピュータが人間の言葉を理解し、処理できるようにする技術

形態素解析・構文解析・意味解析・機械翻訳などがある。

分かち書き処理（形態素解析）

1. ルールベース型

あらかじめ定義された言語の文法ルールに基づいて分かち書きを行う。

2. 統計ベース型

機械学習によって導き出されたルールに基づいて分かち書きを行う。

課題

- ・日本語のあいまいさ
- ・専門用語などへの対応

termextract

- ・専門用語や複合語を抽出できるモジュール
- ・それらの重要度も分析可能

図 4: テキストマイニングと自然言語処理

3.1 特許情報のベクトル化

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

莫大な量の特許情報をベクトル化し、可視化することで特許の技術分野や技術トレンドなどを把握することができると思われる。具体的には特許本文を Sentence-Bidirectional Encoder Representation from Transformer (Sentence-BERT) を用いることで文章全体を単位にベクトル化を行う。

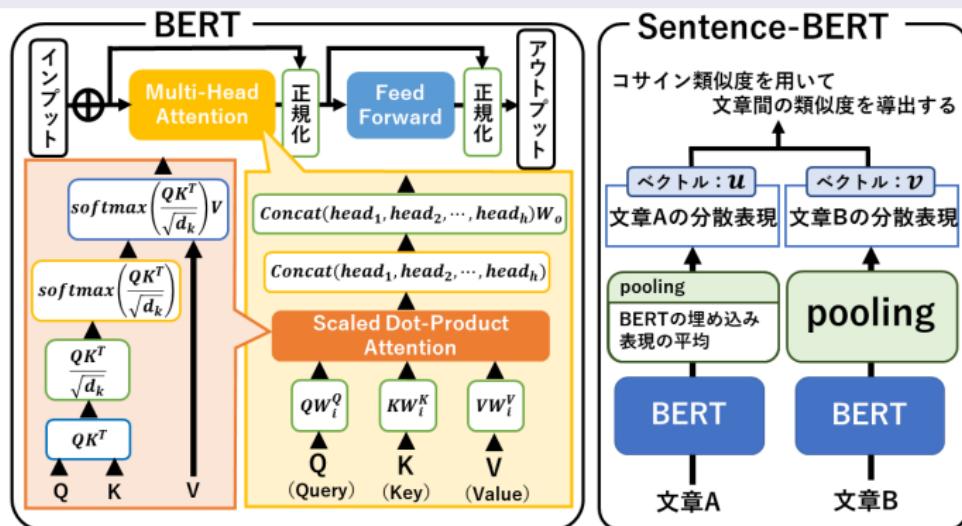


図 5: Sentence-BERT によるベクトル化

3.2 次元圧縮手法とクラスタリング手法

8/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

次元の呪いを回避するため、次元圧縮を行ったのちクラスタリングを行う。ベクトル同士の近さを保有する必要があるため非線形圧縮手法を用いる。非線形圧縮手法には Uniform Manifold Approximation and Projection of Dimension Reduction(UMAP) を用いる。

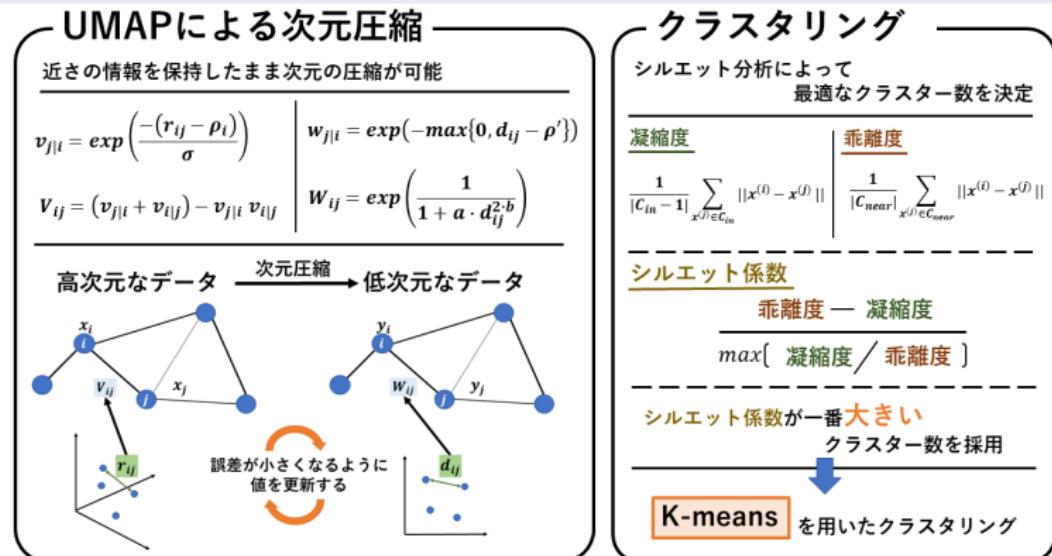


図 6: 次元圧縮とクラスタリング

3.3 単語間のつながりと共起語ネットワーク

9/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

関連性の高い単語は、一緒に出現することが多いため、それらの単語の共起関係を調べることで、単語間の関係性を理解することができる。具体的には共起関係にある単語と単語を線で結んで描かれる共起語ネットワークを作成する。

単語の共起関係

共起分析を行うことで単語間の関係性を導き出す

<p>Jaccard係数</p> <p>共通要素が占める割合が大きければ類似度が高い</p> $J(A, B) = \frac{ A \cap B }{ A \cup B }$ <p><欠点> お互いの要素数の差による影響が大きい</p>	<p>Dice係数</p> <p>Jaccard係数よりも要素差による影響が少ない</p> $DSC(A, B) = \frac{2 A \cap B }{ A + B }$ <p><欠点> お互いの要素数の差による影響がまだ大きい</p>
--	---

Simpson係数

Jaccard係数/Dice係数よりもさらに要素差による影響が少ない

$$overlap(A, B) = \frac{|A \cap B|}{\min\{|A|, |B|\}}$$

<欠点>
片方の要素が少なすぎる → 要素差の許容範囲を設定すると1に近い値になってしまう

共起語ネットワーク

分析結果をもとにネットワークを作成

Simpson係数のしきい値設定

お互いの要素の差の許容範囲を**5000**に設定

係数の値が1になることはめったにない
そもそも一つの単語だったものが分かれて抽出されてしまった可能性が高い

係数の値が**1**になるものは除外

Three.js

Webブラウザ上で3次元コンピュータグラフィックスを描画するためのJavaScriptライブラリ

「3D Force-Directed Graph」というモジュールを用いて3Dでネットワークを作成

図 7: 共起語の分析

4.1 Google Patents からのデータ収集の高速化と分類

10/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

特許プラットフォームである Google Patents から特許をスクレイピングによって収集する。Google Patents では一度に最大 1000 件までしか、表示することができないため、1 年ごとに区切ってスクレイピングを行う。その後分類を行う。



図 8: スクレイピングと分類

4.2 クラスターの解釈と共起語ネットワーク

11/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

表示されているクラスターがどのようなまとまりなのかを明確にするために、クラスターの解釈の提示を行う。共起語ネットワークは大きさを変更できるように、小、中、大の中から選択できるようにしている。

クラスターの解釈

それぞれのクラスターがどのような集まりなのかを提示する

<クラスターのなかでの重要語を分析し、それらを提示する>

懸念点

>クラスターに含まれる特許の数によっては
膨大な時間が必要になる。

解決策

>クラスターの中から**代表的**な特許を取り出し
それらの中から重要語を分析する。

解釈の提示

K-meansによって求められた

クラスターの重心との**ユークリッド距離**が近いものを取り出し
その中で重要語を分析する。

分析によって得られた重要語を

3語表示することで

そのクラスターの解釈とする。

専門用語の抽出と 重要語の分析

- ・Termextractを用いて専門用語と
それらの重要度を取得する。
- ・抽出された専門用語を
Janomeの辞書に登録する。

ユーザー辞書のフォーマット (csv形式)

表層系	左文脈	右文脈	コスト	品詞	分類1	分類2	分類3	活用型	活用形	原型	音読み	発音
-----	-----	-----	-----	----	-----	-----	-----	-----	-----	----	-----	----

共起語ネットワーク

3Dグラフ

>**Three.js**を用いて描画。
大規模なネットワークに適している。

2Dグラフ

>**graphviz**を用いて描画。
小規模なネットワークに適している。

図 9: クラスターの解釈とネットワークの描画

4.3 システム化と IPLへの活用

12/17

動画で紹介

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

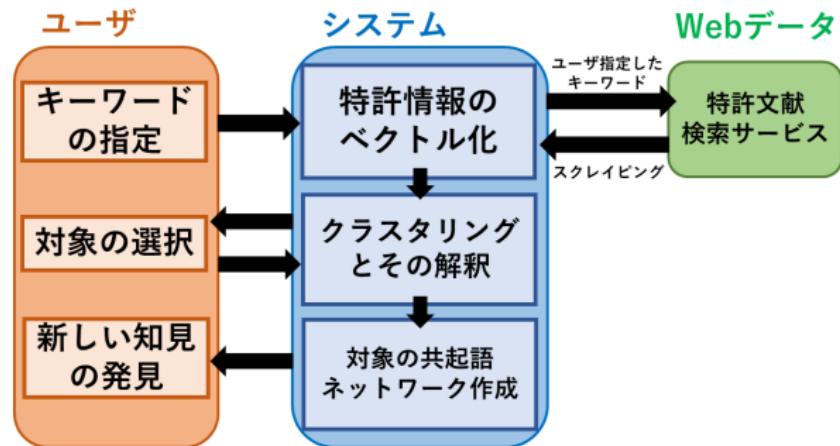


図 10: システムの流れ

5.1 実験の概要

13/17

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

システムの有用性の検証として、

- 実際の事例を設けての実践
- アンケートでの調査

を行う。実際の事例では、「ブロックチェーン技術を持っている会社の、技術の決済システムへの使い道の検討と現状の把握」という題材をもとに実践を行った。

アンケートは5段階のリーカード尺度を用いる。利用者各自に実際にキーワードを考えてもらい、システムを利用してもらった。

アンケート内容

システムの操作性はわかりやすいか	システムの機能は理解しやすいか
レイアウトは親切か	デザインは見やすいか
ストレスなく利用することができたか	クラスターの内容を理解することができたか
共起語ネットワークによる出力は適していたか	3Dグラフによる提示は適切であるか
効率的な特許探索を行えると思ったか	新しい知見を発見できそうか

5.2 実験結果と考察

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

「コンサート」や「グッズ」などから「ファン通貨」という単語につながりがあることから、アーティストのファン特有の通貨をブロックチェーン技術を用いて作り出し、ファンのコミュニティー内でその通貨を発行することが考えられる。

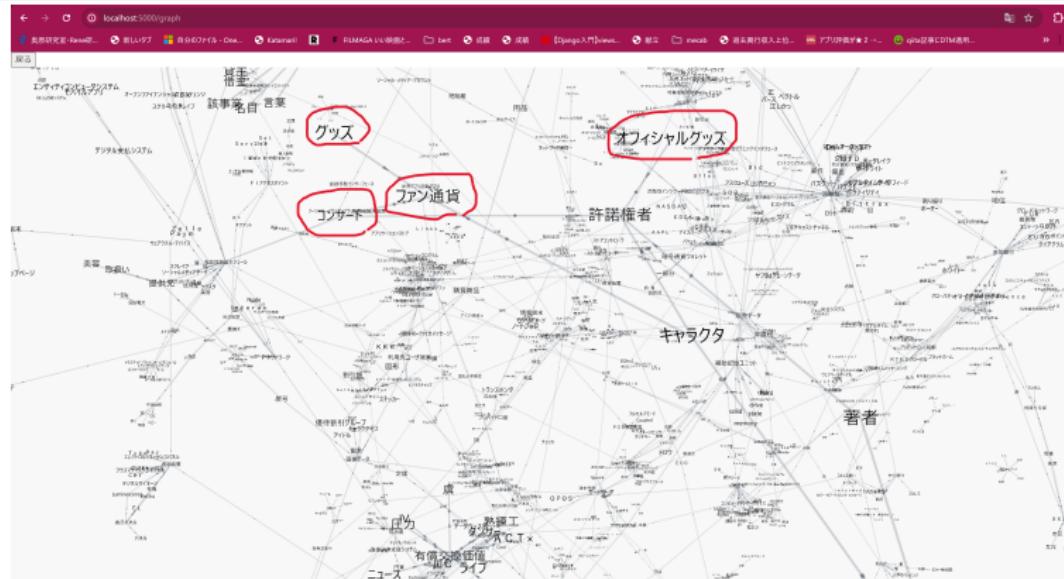


図 11: 作成された 3D グラフ

5.2 実験結果と考察

アンケート結果を以下に示す。

表 1: アンケート結果

	回答者A	回答者B	回答者C	回答者D	回答者E
システムの操作性はわかりやすいか	4	4	5	4	4
システムの機能は理解しやすいか	3	5	4	5	4
レイアウトは適切か	4	4	5	4	5
デザインは見やすいか	5	4	4	5	5
ストレスなく利用することができたか	2	2	3	2	2
クラスターの提示は適切であるか	4	4	2	3	4
共起語ネットワークは適切であるか	3	4	5	2	5
3Dグラフによる出力は適切であるか	3	4	4	5	5
効率な特許探索を行えそうか	5	4	5	4	4
新しい知見を発見できそうか	4	5	5	4	4
入力してもらったキーワード	・スマホ ・キーホル ダー	・アジ ・餌	・ネット ワーク ・アローダ イヤグラム	・音楽 ・楽曲 ・ゲーム	・アメリカ ・インド ・ドイツ

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

5.2 実験と考察

16/17

アンケート結果の考察

Q.5 ストレスなく利用することができたか



この項目においてあまりいい評価が得られなかった理由として

- 入力するキーワードによって取得する特許の数が異なり、多い場合に最小年数の6年であっても時間がかかることがある。
- 共起語ネットワークを作成する際に行う分かち書き処理においてクラス名に含まれる特許の数が多いと処理に時間がかかることがある。
- 3Dグラフの描画が遅くノードを動かすと動作が重くなってしまう。

などがあげられる

Q.9/10 効率的な特許探索を行えそうか/新たな知見を得ることができそうか

この項目においては全体的にいい評価を得ることができた。

○ 知見を得るために特許探索を行えるシステムとして有用であるといえる。

6. おわりに

17/17

まとめ

1. はじめに
2. 知的財産戦略と特許情報処理
3. 特許情報の可視化
4. 提案手法
5. 実験結果並びに考察
6. おわりに

今後の課題