

Visualization of Research Topics Using Word Cloud Technology- Characteristics of Research Trends in Article Titles and Keywords -

ワードクラウド技術を用いた研究トピックスの可視化

— 論文タイトル・キーワードにおける研究トレンドの特徴 —

小笠原 功明 [Naruaki Ogasawara]

一般社団法人 日本内科学会 編集部

責任著者：小笠原 功明

Email : n-ogasawara@naika.or.jp

Key words : 学術情報, 学術流通, 学会誌, 学術誌, 学術出版, テキストマイニング手法, N-gram 解析, 論文タイトル分析, 著者キーワード分析, ワードクラウド, 研究トレンド, 医学, 内科学

【抄録】

日本内科学会が発行する「Internal Medicine」は、2022年にImpact Factorが過去最高を記録した。その要因を探るべく、直近10年間(2014~2023年)に掲載された6,868本の論文を対象に、論文タイトルおよび著者キーワードをテキストマイニングによってN-gram解析を行った。加えて、ワードクラウドにより、直近10年間の研究トレンドの可視化を行った結果、研究トレンドの把握に、著者キーワードの分析および可視化が有用なことが確かめられたので、ここに紹介させていただく。

1. イントロダクション - はじめに -

日本内科学会では、「日本内科学会雑誌」(和文誌)と「Internal Medicine」(英文誌)を発行している¹⁾。和文誌が、解説記事(Editorial)や各分野の専門家による総説論文(トピックス)を中心とする雑誌であるのに対し、英文誌は、総説論文(Review Article), 原著論文(Original Article), 症例報告(Case Report), 画像による症例報告(Pictures in Clinical Medicine), エディターへの手紙(Letters to the Editor)ならびに論説(Editorial)を掲載している²⁾。

2. 英文誌「Internal Medicine」の役割

日本内科学会が発行する英文誌「Internal Medicine」は、内科学分野における研究成果を発信する重要な媒体として位置付けられており³⁾,

研究成果を公表できる場となっている。

同誌は、学会誌の質と影響力を示す指標であるImpact Factor (IF) を取得しており、質の高い研究成果を掲載することで、有益な情報を、より多くの研究者に提供している。

3. Impact Factor (IF) の推移

Internal Medicine 誌は、症例報告を主体とした雑誌ではあるが、2022年にImpact Factorが過去最高を記録している。

前項のとおり、IFが過去最高値を記録した事実は、雑誌そのものが、質の高い臨床経験や研究を紹介し、広範囲の研究者により引用されていることにほかならない。

そこで、本稿では、直近10年間に掲載された論文のタイトルと著者キーワードをテキストマイ

ニング手法とワードクラウド技術によって分析を行い、掲載された論文の特徴や研究トピックスについて可視化を行ってみたい。

4. 直近 10 年間に掲載された論文の分析 (2014~2023 年に掲載された 6,868 本対象)

直近 10 年間(2014~2023 年)において、Internal Medicine 誌に掲載された論文タイトル 6,868 本を対象に、下記の手順と解析環境で、テキストマイニング手法を用いて「可視化分析」を行った。

1) 分析手順と解析環境

【分析手順】

- ① 論文タイトルのデータ取得 (6,868 件)
- ② 論文に紐づく著者キーワードを取得
- ③ 自然言語処理 (単語の分割と小文字化)
- ③ 不要語 (Stop Word) の除去処理
- ④ 頻出単語の抽出
- ⑤ 自然言語処理 (N-gram 処理)
- ⑥ ワードクラウド生成

【解析環境】

- ・ OS / Windows 10 Pro 22H2
- ・ 解析言語 / Python 3.10.5
- ・ データ解析ライブラリ / Pandas 2.2.2
- ・ 言語処理ライブラリ / NLTK 3.8.1
- ・ グラフ描画ライブラリ / Matplotlib 3.8.4
- ・ 可視化ライブラリ / Wordcloud 1.9.3
- ・ 開発環境 (IDE) / PyCharm

(2024.4.17 解析実施)

3) 解析結果

テキストマイニング手法を用いた「可視化分析」結果は、図 1~図 6 のとおりである。

文字の大きさは出現回数の多少を表し、色の濃さは、他の単語と比較した際の単語の頻度の高さを表している。

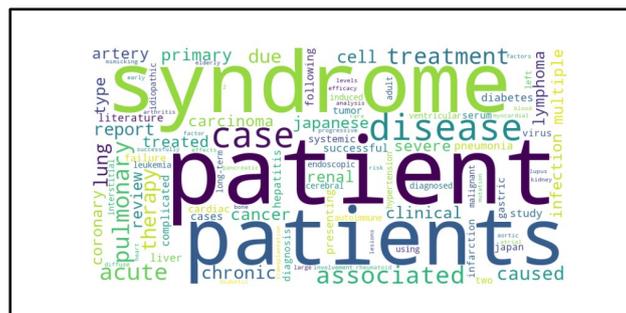


図 1. 論文タイトル (N=1) のワードクラウド

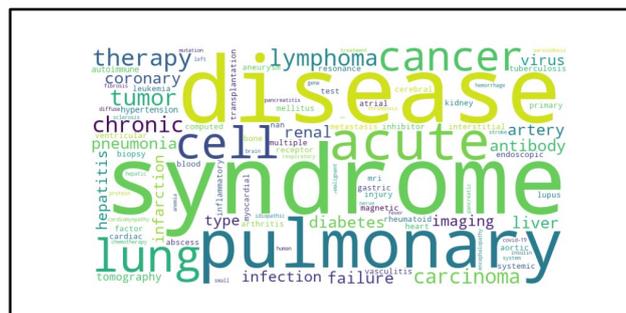


図 2. 著者キーワード (N=1) のワードクラウド

N-gram が 1 の場合は、論文タイトルには、patient(1,757), syndrome(1,537), patients(1,423) など、一般的な単語が多く含まれていた。著者キーワードには、syndrome(1,918), disease(1,763), pulmonary(830), acute(774), lung(699) など、具体的な臓器名 (肺) や特定の症候群、疾患名、治療を想起させるものが多く含まれていた。

同様に、N-gram の値を増やした場合について、論文タイトルと著者キーワードの違いを検証したい。

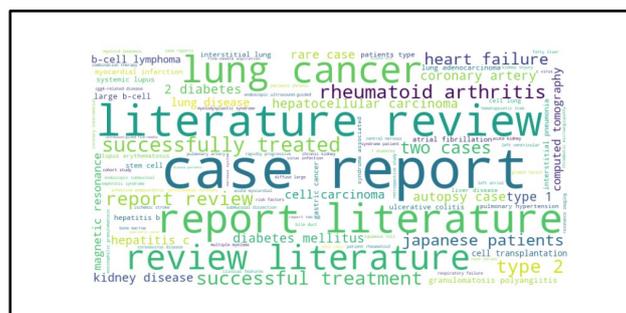


図 3. 論文タイトル (N=2) のワードクラウド

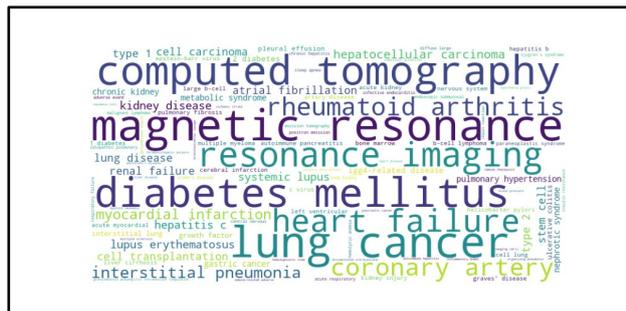


図4. 著者キーワード (N=2) のワードクラウド

N-gram が 2 の場合は、論文タイトルには、上位に、case report (296), literature review (174), report literature(147)など、論文種別や研究手法に関する単語が多く含まれていた。

著者キーワードには、magnetic resonance (287), lung cancer(281), computed tomography (251), diabetes mellitus(249), heart failure (198), rheumatoid arthritis(172)など、具体的な症候名、疾患名、検査方法に関するが多く含まれていた。

N-gram が 3 の場合は、論文タイトルには、上位に、case report literature(146), report literature review(146), case report review (81), report review literature(76)など、論文種別(症例報告)に関する単語が多く含まれていた一方、具体的な疾患名として、type 2 diabetes(67), large b-cell lymphoma(43), interstitial lung disease(39), systemic lupus erythematosus(37), type1 diabetes (32), hepatitis c virus(32)などが多く含まれていた。

著者キーワードには、magnetic resonance imaging (234), systemic lupus erythematosus (121), stem cell transplantation (96), chronic kidney disease (89), type 2 diabetes (88), interstitial lung disease (75), acute myocardial infarction (70), acute kidney injury (66), large b-cell lymphoma(65), type 1 diabetes (64)など、具体的な疾患名や検査方法に関する単語が多く含まれていた。



図5. 論文タイトル (N=3) のワードクラウド

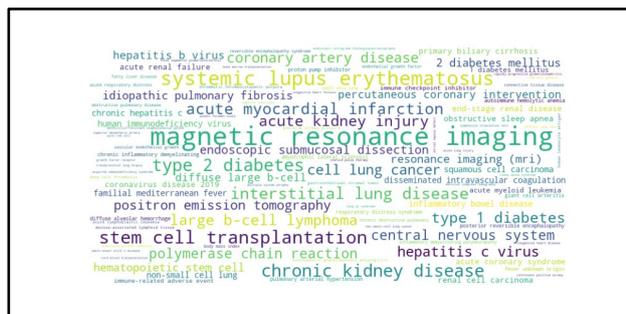


図6. 著者キーワード (N=3) のワードクラウド

4. まとめ

今回、Impact Factor が過去最高を記録した要因を探るべく、直近 10 年間 (2014~2023 年) に掲載された論文タイトル、著者が設定したキーワード(著者キーワード)を対象に、テキストマイニング手法とワードクラウド技術によって「可視化分析」を行った。

結果として、N-gram の値 (N=x) が大きくなると、より具体的な疾患名や医学用語が抽出されることがわかった。加えて、著者キーワードには、論文タイトルに比べて、より具体的な疾患名や医学用語(検査方法等)が含まれていることがわかった。

以上により、ワードクラウドを用いて可視化することで、研究トピックスを把握する際に、有益な情報を得る可能性があることがわかった。

(抄録 : 201 単語, 本文 : 1,824 単語)

——— 参考文献 ———

- 1) 小笠原功明. 日本内科学会英文論文誌「Internal Medicine」の取り組み. 情報管理 60 (5), 316-324, 2017
- 2) Naruaki Ogasawara, Kazuto Matsunaga, Hajime Isomoto et al. Internal Medicine Year in Review 2022. Intern Med 62 23:3431-3435, 2023
- 3) 清水渉. 日本内科学会英文論文誌「Internal Medicine」のオープンアクセス化について. 日本内科学会 創立 120 周年記念誌 P. 105-106, 2023