




オープンサイエンス賞(テーマ賞)

プレスリリース論文の可視化

<https://k2taro.github.io/PressReleaseVis/>



2022/3/13

LODチャレンジ2021 授賞式

大学や研究機関のプレスリリース

研究成果として有名な学術誌に論文掲載された内容を取り上げることが多い。レベルの高い研究成果＋日本語紹介(プレスリリース自体のこと)もあるためちょうど良い題材と考えた。

本作品では以下の2研究機関を対象とした。

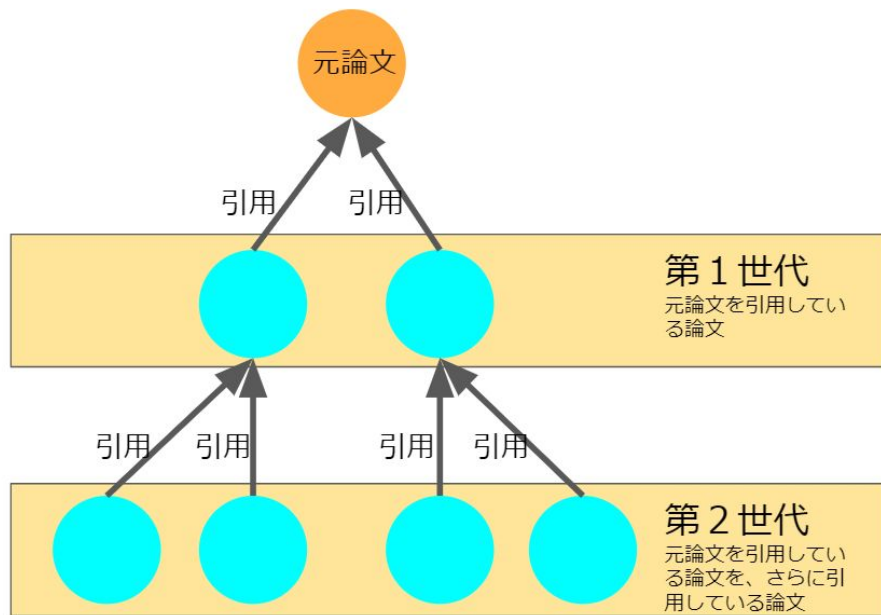
- 科学技術振興機構*₁ 2018年度*₃
- 理化学研究所*₂ 2018年*₃

*₁:大学等に研究資金を配分する機関。配分先の大学等と連名でプレスリリースを行う。

*₂:自然科学系全般を扱う国立の研究所

*₃:論文が引用されるにはある程度の時間経過が必要なため～3年前とした。

何をやっているのか？



*:引用している論文の情報は、[COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations](#)から取得

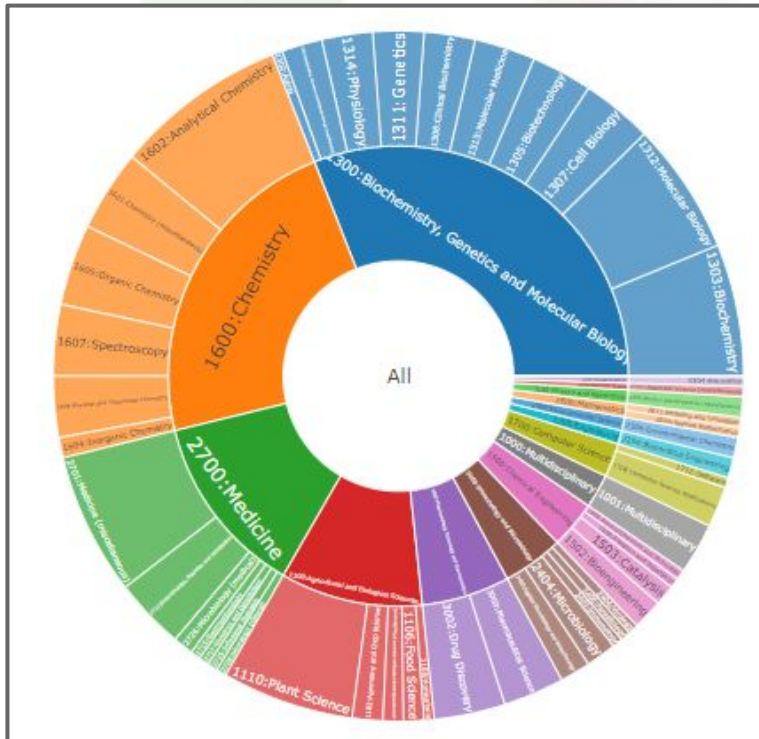
プレスリリースの成果論文を引用している論文の情報(第1世代)、引用している論文をさらに引用している論文(第2世代)について分析・可視化

こんなことがわかるかも？

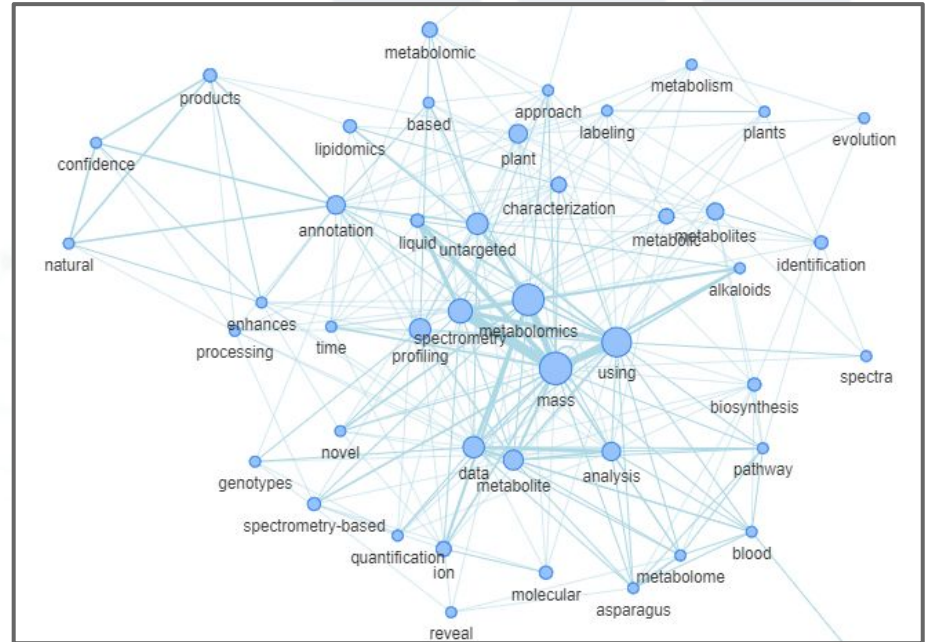
- どのような論文から引用されているのか？
- 第1世代->第2世代で著者の広がりが増えていく？
- 第1世代->第2世代で分野のひろがりが増えていく？

可視化分析の例

掲載誌の分野の可視化



タイトルに含まれる単語の可視化



単語	頻度
mass	19
metabolomics	18
using	16
spectrometry	11
data	9

オープンサイエンス賞をもらったので...

- 主なデータソースであるOpenCitationsは、これ自体がオープンサイエンスを目的とした取り組み*
- 他のさまざまな団体や学術出版社を巻き込んだI4OCというイニシアティブの成果でもある*
- 引用(citation)のオープンデータセットを作る目的
 - プロプライエタリなデータが主流で、コストを支払える一部の者しかアクセスできない
 - データセットがオープンでないため再現性の担保が難しい(第三者による検証が困難) etc...

(そもそも、こんなに素晴らしいデータセットがあるので「使わないともったいない」がモチベーションでした)

*Silvio Peroni, David Shotton; OpenCitations, an infrastructure organization for open scholarship. Quantitative Science Studies 2020; 1 (1): 428–444. doi: https://doi.org/10.1162/qss_a_00023